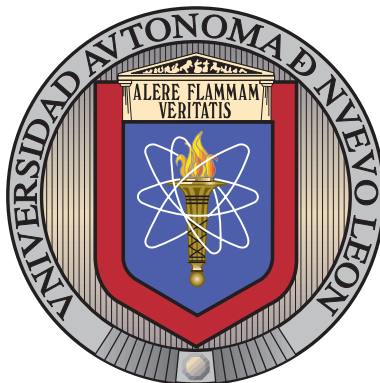


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECTOR DE ESTUDIOS DE POSGRADO



AGRUPAMIENTO GEOGRÁFICO DE LA DEMANDA
PARA LA PLANEACIÓN DEL TRANSPORTE Y
DISTRIBUCIÓN

POR

ALAN EMILIO PAHUA REYES

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
MAESTRÍA EN LOGÍSTICA Y CADENA DE SUMINISTRO

ENERO 2020

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECTOR DE ESTUDIOS DE POSGRADO



AGRUPAMIENTO GEOGRÁFICO DE LA DEMANDA
PARA LA PLANEACIÓN DEL TRANSPORTE Y
DISTRIBUCIÓN

POR

ALAN EMILIO PAHUA REYES

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
MAESTRÍA EN LOGÍSTICA Y CADENA DE SUMINISTRO

ENERO 2020


Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica
Subdirección de Estudios de Posgrado

Los miembros del Comité de Tesis recomendamos que la Tesis «Agrupamiento geográfico de la demanda para la planeación del transporte y distribución», realizada por el alumno Alan Emilio Pahua Reyes, con número de matrícula 1937571, sea aceptada para su defensa como requisito parcial para obtener el grado de Maestría en Logística y Cadena de Suministro


El Comité de Tesis



Dr. Miguel Mata Pérez
Asesor

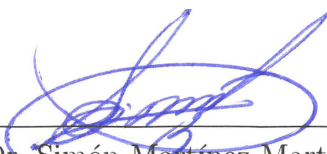


M.C. Carlos Alberto Álvarez Herrera
Revisor




Dr. Luis Alfonso Infante Rivera
Revisor

Vo. Bo.



Dr. Simón Martínez Martínez
Subdirector de Estudios de Posgrado



San Nicolás de los Garza, Nuevo León, enero 2020

Dedicó está tesis a Dios, a mi familia y a mi mascota Odi

ÍNDICE GENERAL

Agradecimientos	xii
Resumen	xiii
1. Introducción	1
1.1. Planteamiento del problema	3
1.2. Objetivo	4
1.3. Hipótesis	4
1.4. Justificación	4
1.5. Metodología	5
1.6. Estructura de la tesis	6
2. Antecedentes	8
2.1. El transporte en la cadena de suministro	8
2.1.1. Regionalización	10
2.1.2. Accesibilidad a regiones	12
2.2. Agrupamiento (<i>Clustering</i>)	14

2.2.1. Algoritmos de agrupamiento	17
2.2.2. Algoritmos de jerárquicos	19
2.2.3. Algoritmos particionales	20
2.2.4. Algoritmo k -medias	21
2.2.5. Componentes de una tarea de agrupamiento	22
2.2.6. Validación de los grupos	24
2.3. Aplicaciones de algoritmos de agrupamiento en logística	27
2.4. La importancia de los pronósticos en la planeación	33
2.4.1. Clasificación de los métodos de predicción	35
2.4.2. Selección del método de pronóstico	39
2.4.3. Medición del error de un método de pronóstico	40
2.4.4. Conclusión del capítulo	43
3. Metodología	44
3.1. Descripción general	44
3.2. Fases preliminares	45
3.3. Fase 1: Análisis de la demanda	46
3.4. Fase 2: Selección de la técnica de agrupamiento	47
3.5. Fase 3: Validación de resultados	48
3.6. Fase 4: Diseño de clústeres	48
3.7. Fase 5: Predicción	49

3.8. Conclusiones del capítulo	50
4. Implementación y análisis de resultados	51
4.1. Fases preliminares	51
4.1.1. Identificación de áreas de oportunidad y definición del problema	52
4.1.2. Revisión sistemática de la literatura (SRL)	54
4.2. Implementación de la fase 1: Análisis de demanda	57
4.2.1. Extracción de características	58
4.3. Resultados de la fase 2: Selección del algoritmo de agrupamiento . . .	61
4.3.1. Aplicación del algoritmo	62
4.4. Implementación de la fase 3: Validación de resultados	66
4.5. Implementación de la fase 4: Diseño de clústeres	66
4.6. Implementación de la fase 5: Predicción	67
4.6.1. Determinación del esquema de la serie temporal	69
4.7. Experimentación con modelos de predicción	71
4.7.1. Modelo por descomposición	71
4.7.2. Experimentación con el modelo ARIMA	72
4.7.3. Experimentación con el modelo Holt Winters	74
4.7.4. Análisis comparativo	77
4.7.5. Ejecución del modelo Holt Winters a los clústeres resultantes .	79
4.7.6. Análisis de sensibilidad	80

4.8. Conclusiones del capítulo	81
5. Conclusiones	83
5.1. Conclusiones	83
5.2. Contribuciones	84
5.3. Recomendaciones	85
5.4. Posibles líneas de investigación a futuro	85

ÍNDICE DE FIGURAS

1.1. Componentes del costo logístico total	2
2.1. Regiones económicas de México	11
2.2. Clasificación de carreteras	12
2.3. Agrupamiento (<i>Clustering</i>)	16
2.4. Características del clúster	17
2.5. Dendrograma, AHC, k=7	20
2.6. Patrones de comportamiento	39
3.1. Esquema metodológico	45
3.2. Revisión sistemática de literatura	46
3.3. Clasificación de los modelos de predicción	49
4.1. Enfoques de agrupamiento	56
4.2. Métodos de agrupamiento	61
4.3. Evolución de varianzas	63
4.4. Clases (clústeres) resultantes	64

4.5. Formación de regiones	67
4.6. Panel de análisis para base de datos	68
4.7. Demanda histórica acumulada	69
4.8. Tipo esquema para la serie temporal	71
4.9. Modelo por descomposición	72
4.10. Modelos ARIMA evaluados	73
4.11. Modelo Holt Winters ajustado	77

ÍNDICE DE TABLAS

4.1. Portales literarios	55
4.2. Variables de agrupamiento	59
4.3. Estructura tabular (a)	59
4.4. Estructura tabular (b)	59
4.5. Evolución de varianzas	62
4.6. Agrupamiento de los clientes según k -medias	65
4.7. Análisis descriptivo del esquema temporal	70
4.8. Modelos ARIMA $(p,d,q)(P,D,Q)$	74
4.9. Análisis comparativo de los modelos predictivos evaluados	78
4.10. Análisis de error en el pronóstico a nivel clúster	80
4.11. Análisis de sensibiidad basado en los tres escenarios	81
4.12. Análisis de sensibiidad basado en los tres escenarios	81

AGRADECIMIENTOS

Agradezco hoy y siempre a mis padres Juan Manuel Pahua Molina y Maricela Reyes Ambris por brindarme su apoyo su amor, su trabajo y su confianza absoluta.

A mis hermanos, Daniela y Juan Luis, por su ejemplo, admiración y motivación constante por seguir adelante. También agradezco a mi mascota Odi, por contagiarme siempre su alegría al recibirme.

A mis abuelos por su fortaleza y perseverancia.

Agradezco al comité de tesis asignado a este trabajo de investigación, al Dr. Miguel Mata Pérez, por haber sido mi tutor y amigo; por el apoyo y el tiempo dedicado, así como su experiencia y conocimientos compartidos. Así también, al M.C. Carlos Álvarez, por abrirme las puertas de la empresa caso de estudio y creer en mi potencial. De igual manera agradezco al Dr. Luis Infante por sus conocimientos, su asesoría y su gran disposición.

Finalmente, agradezco a los profesores de la Maestría en Logística y Cadena de Suministro, a la Universidad Autónoma de Nuevo León (UANL), a la facultad de Ingeniería Mecánica y Eléctrica (FIME), por supuesto a la empresa caso de estudio y por último, pero no menos importante a CONACyT por la beca otorgada en mis estudios de posgrado.

RESUMEN

Alan Emilio Pahua Reyes.

Candidato para obtener el grado de Maestría en Logística y Cadena de Suministro .

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: AGRUPAMIENTO GEOGRÁFICO DE LA DEMANDA PARA LA PLANEACIÓN DEL TRANSPORTE Y DISTRIBUCIÓN.

Número de páginas: 95.

OBJETIVOS Y MÉTODO DE ESTUDIO: El estudio está motivado por un problema real de una empresa internacional instalada al noreste del país. El objetivo de la investigación es: predecir la demanda de transporte para generar planes tácticos y financieros efectivos y anticipados por medio de herramientas de agrupamiento y predicción.

La metodología está basada en la literatura especializada y ha sido adaptada para nuestro caso de estudio. Comprende cinco fases, la primera fase consiste en el tratamiento de datos y la selección de las variables que serán empleadas para la clusterización, la segunda fase se centra en la elección del algoritmo de agrupamiento, la tercera fase trata sobre la validación de los grupos sugeridos por el algoritmo de agrupamiento, la cuarta fase se enfoca en el diseño de clústeres, finalmente la quinta

fase consiste en la construcción y aplicación de modelos de predicción.

CONTRIBUCIONES Y CONCLUSIONES: Nuestra contribución principal a la empresa caso de estudio fue entonces, crear, aplicar y validar una metodología basada en herramientas cuantitativas y una serie de recomendaciones técnicas para la elaboración de planes tácticos y toma de decisiones efectiva en el de área de transporte y distribución.

Concluimos que haciendo uso de herramientas cuantitativas (agrupamiento y predicción) queda demostrado que en la implementación ya sea en conjunto o por separado de las herramientas abordadas, se obtiene en promedio una mejora del 10 % en el proceso de planificación del transporte.

Firma del asesor: _____

Dr. Miguel Mata Pérez

CAPÍTULO 1

INTRODUCCIÓN

El transporte es un componente vital y de alto interés para las organizaciones como proceso para el traslado oportuno de productos al mercado, motivo que lo convierte en una operación logística clave en el desarrollo de las actividades vinculadas con la planeación, abastecimiento, movimiento de los bienes, flujos de información y satisfacción del cliente (Arreola *et al.*, 2013; Corredera, 2012; Cendrero y Truyols, 2008).

Según Long (2006) y Hurtado *et al.* (2018), el transporte merece ser estudiado por representar millones de dólares gastados en ferrocarriles, aerolíneas, camiones y barcos ya que economía global está basada en la capacidad de transportar bienes alrededor del mundo. La figura 1.1 muestra como en la cadena de suministro cerca del 49 % del costo logístico de una empresa es absorbido por el transporte (Davis *et al.*, 2011), también es comúnmente aceptado que los costos de transporte normalmente se hallan entre un tercio y dos tercios de los costos logísticos totales (Ballou, 2007).

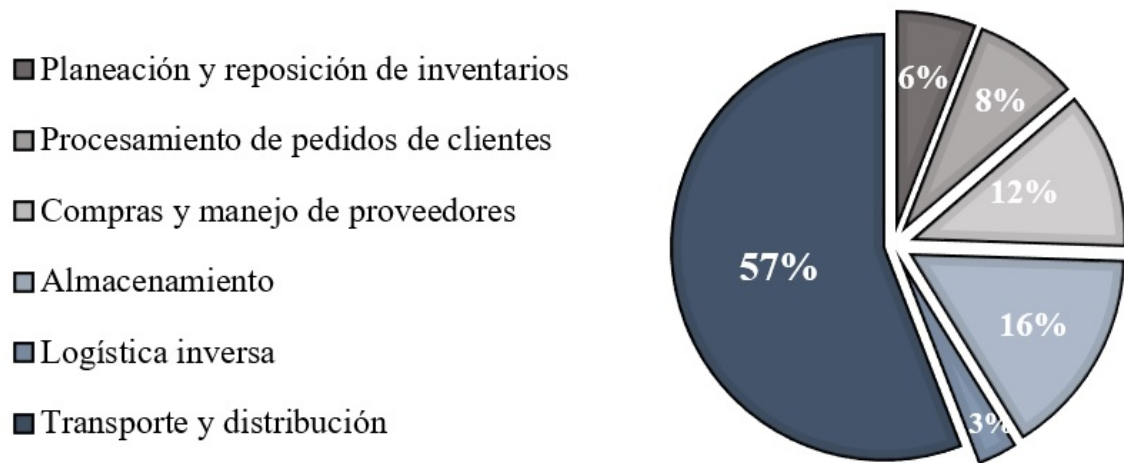


Figura 1.1: Componentes del costo logístico total

Fuente: Elaboración propia basada en el caso del estudio y soportada Ballou (2007); Davis *et al.* (2011); Arreola *et al.* (2013).

De acuerdo a lo anterior, la planeación del transporte es una tarea que debe anticiparse haciendo uso de modelos de predicción que permitan estimar el número de cajas (vehículos de carga) que se emplearán para movilizar los bienes y por supuesto establecer los planes financieros que conlleva la actividad de movilizar la mercancía (Lu y Kao, 2016).

El proceso de transporte y distribución es comúnmente atendido por la comunidad científica y empresarial por significar el principal costo de los costos logísticos totales. Se presenta el caso de una empresa dedicada a la manufactura y distribución de productos terminados cuyo control del abanico de clientes y proceso de planeación del transporte son actividades complejas. Para ello se propone el uso de herramientas matemáticas que permitan realizar predicciones de la demanda del transporte basadas en el diseño de grupos con características estadísticamente similares dentro de estos y disimilares entre los mismos. Para realizar las predicciones de la demanda del transporte de los clústeres resultantes, se experimenta con modelos de pronóstico tradicionales tales como método por descomposición, método Holt Winters y el modelo Box-Jenkins o ARIMA (*autoregressive integrated moving average*), identificando el modelo que mejor se ajuste al comportamiento real de la demanda de flotillas. En

este documento se describe la metodología que servirá como base para la aplicación del diseño de clústeres, pronósticos en la demanda del transporte, así también para posteriores aplicaciones de planeación hacia los tres horizontes (estratégica, táctica y operativa) y diseño de rutas de distribución.

1.1 PLANTEAMIENTO DEL PROBLEMA

La presente tesis aborda un problema del entorno real de una compañía dedicada a la fabricación y comercialización de bienes manufacturados instalada al noreste de México. La empresa tiene operaciones en todo el país por lo que la capacidad de respuesta que el mercado exige es inmediata, donde anticipar las necesidades de la empresa y de sus clientes se convierte en una actividad clave para solucionar los desafíos estratégicos y operativos.

De acuerdo con Lightner *et al.* (2016) las empresas dedicadas al giro manufacturero y de servicios tienen complicaciones con la planeación del transporte y la distribución de sus productos debido a la falta de aplicación de herramientas técnicas en: (1) la proyección de la demanda, (2) la ubicación de los depósitos, (3) el control de la cartera de clientes, (4) la gestión de flotas, (5) el aumento del nivel de servicio y (6) la disponibilidad de productos. En esta investigación se tiene particular interés en trabajar sobre el control de clientes y proyección de la demanda de vehículos de carga para apoyar la toma de decisiones en el proceso de planeación para la logística de salida del caso de estudio.

El extenso abanico de clientes que maneja la empresa dificulta el proceso de planeación aguas arriba, así como aguas abajo para el transporte y distribución. Por consecuencia los costos asociados a la contratación de flotas se elevan por motivos de renta de camiones inmediata, se presenta falta de disponibilidad de transporte que satisfagan las rutas de distribución y ocurre desalineación con el departamento de finanzas, quienes liberan el presupuesto acordado una vez recibida la estimación

de recursos que el área de logística requerirá. Eventualmente el nivel de servicio al cliente disminuye puesto que las ordenes completadas estan por debajo del total de órdenes del sistema. Dado un conjunto (n) de clientes la empresa desea dividirlos en un número determinado de (k) clústeres o agrupaciones de acuerdo criterios tales como volumen de ventas, demanda de transporte, tipo de vehículo, capacidad de vehículo, distancia, tiempo de entrega y accesibilidad que son variables comúnmente usadas en análisis de grupos (Díaz *et al.*, 2010; Dondo y Cerdá, 2007).

1.2 OBJETIVO

Pronósticar la demanda de transporte para generar planes estratégicos efectivos y anticipados que permitan ahorrar recursos financieros destinados a operaciones de transporte y distribución por medio de herramientas de agrupamiento y predicción.

1.3 HIPÓTESIS

Por medio de herramientas matemáticas de clasificación y predicción es posible realizar planes estratégicos y financieros confiables a partir de la predicción efectiva de la demanda de transporte para clientes con características homogéneas.

1.4 JUSTIFICACIÓN

Los costos logísticos representan entre 18 % y 35 % del valor del producto final, cifra que supera la que registran los países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), que es de 8 %. En relación al porcentaje del producto interno bruto, los costos logísticos en América Latina y el Caribe están en un rango superior de entre 50 % y 100 % en comparación a los países

de la OCDE (BID, 2013). Por ello, uno de los principales motivos de la empresa caso de estudio para medir el desempeño logístico es reducir los costos de la logística iniciando por los costos de transporte los cuales representan aproximadamente el 50 % de las actividades logísticas (Sheu, 2007). La estimación de unidades motrices para actividades de carga y distribución y la clasificación de clientes, son dos de los problemas principales a los que se enfrentan los analistas y planeadores de la compañía en cuestión durante la reunión mensual de S&OP (*sales and operations Planning*) , donde típicamente los tomadores de decisiones suelen modificar los planes estratégicos y operativos en base a su experiencia (Lu y Kao, 2016). Las predicciones basadas en técnicas de agrupamiento reflejan beneficios logísticos significativos analizados en reuniones con la empresa caso de estudio y vistos en la literatura, además de reducir costos de transporte y distribución, se tienen importantes ventajas que motivan a este trabajo de tesis, tales como: generar una base para la toma de decisiones hacia los tres horizontes de planeación (operativo, táctico y estratégico), controlar del abanico de clientes, alinear estratégica (control de envíos), aumentar la ventaja competitiva de la empresa, incrementar la calidad del servicio, inicializar problemas de optimización en modelación matemática y desarrollar métodos estadísticos entre los cuales destacan modelos de predicción (Shaub, 2019).

1.5 METODOLOGÍA

Hay un consenso respecto a las inquietudes logísticas (planeación del transporte, control de clientes, ruteo de vehículos, localización de instalaciones, inventarios y almacenaje) que son comúnmente estudiadas en la literatura y que acertadamente la empresa de estudio presenta. Identificada esta relación teórica y real se propone una metodología basada en la literatura especializada con sustento en herramientas cuantitativas de análisis para la resolución de la problemática anteriormente presentada. La metodología comprende las siguientes 6 fases: (1) reconocimiento del problema y áreas de oportunidad, (2) revisión bibliográfica, (3) análisis de la demanda, (4) di-

seño de clústeres, (5) predicción de la demanda, (6) resultados y conclusiones. Dichas fases son desglosadas a lo largo del actual documento de investigación.

1.6 ESTRUCTURA DE LA TESIS

El presente documento de tesis está conformado por seis capítulos, los cuales abarcan desde el conocimiento del problema hasta conclusiones, se dividen de la siguiente manera:

Capítulo 1. Introducción: Se presenta la introducción a la investigación realizada y un contexto general sobre ella, ofreciendo la descripción del problema, justificación, objetivo, e hipótesis que guiarán el proyecto de tesis, así como la orientación al lector acerca de cómo se estructura la información a lo largo del documento.

Capítulo 2. Antecedentes: Se presentará la revisión de la literatura, se da una visión general de los aspectos y métodos más relevantes para el agrupamiento de clientes y para pronósticos de la demanda, así como investigaciones recientes de común problemática al presente trabajo de tesis.

Capítulo 3. Metodología: Se muestra detalladamente la secuencia a seguir para el cumplimiento de los objetivos, también se identifican y describen las herramientas cuantitativas empleadas a lo largo de los últimos años que mejor se adapten al problema bajo estudio. La metodología está basada en la recopilación de bibliografía especializada y en las necesidades de la empresa caso de estudio.

Capítulo 4. Implementación y Análisis de resultados: Se aplican las herramientas cuantitativas seleccionadas en capítulos anteriores a través de software especializado, haciendo uso de información proporcionada por el caso de estudio. Por otro lado, se compara la funcionalidad y efectividad de las herramientas matemáticas aplicadas contra información acontecida en la empresa para validar resultados.

Capítulo 5. Conclusiones: Se da a conocer las contribuciones generadas a la

empresa en estudio, también conclusiones y recomendaciones para trabajos futuros

Una vez expuesto este capítulo acerca del contexto general del tema, problema a tratar y de la importancia de la investigación. Los siguientes capítulos están dirigidos al desarrollo teórico y práctico de lo antes descrito, a su vez, se abordan procedimientos de gran interés en estadística y optimización para la cadena de suministro.

CAPÍTULO 2

ANTECEDENTES

El presente capítulo describe los conceptos básicos de la investigación, comenzando con temas concernientes al transporte y el papel que desempeña en logística y cadena de suministro, su relación con la regionalización del país, así como generalidades normativas acorde a la clasificación de carreteras en México. Por otra parte, se mencionan los estudios y herramientas aplicadas al agrupamiento de clientes con fines estratégicos y operativos que han surgido a través del tiempo, y que han sido empleadas como técnicas de arranque para el diseño y análisis de la cadena de suministro. Finalmente, son descritas las generalidades de los modelos predictivos y algunas de sus aplicaciones en la planeación dentro de la cadena de suministro. Por lo antes mencionado este capítulo constituye aspectos clave para el desarrollo y la comprensión del proyecto de tesis.

2.1 EL TRANSPORTE EN LA CADENA DE SUMINISTRO

Dentro de la cadena de suministro se denomina transporte a un sistema formado por múltiples elementos; infraestructura, el vehículo y la empresa de servicio. Estos elementos están interrelacionados entre sí, pues ninguno es útil sin que los otros existiesen (Cendrero y Truyols, 2008).

El transporte pertenece a un sistema logístico, basado en la literatura revisada un sistema logístico es definido como: *“la red de unidades autónomas y coordinadas que permiten garantizar la satisfacción de los clientes finales en cantidad, calidad, tiempo y costos demandados”* (Corredera, 2012).

El transporte juega un papel fundamental en la logística local e internacional, se ocupa de todas las actividades relacionadas con situar los productos en los puntos de destino correspondientes, mediante los criterios de seguridad, costo, calidad del servicio, seguros de la empresa transportista y la entrega de la mercancía (Antún, 2005).

La función del transporte se concreta a llevar los productos en tiempo y forma al sitio donde son requeridos. La interacción del transporte con los programas de abastecimiento y distribución, constituye un proceso dinámico que en la cadena de suministro exige una alta coordinación (IMT, 2002).

Por lo anterior, en la cadena de suministro, la actividad de transporte debe ser planificada y considerada en los acuerdos de coordinación entre los componentes de la misma. A su vez, la logística del transporte en la cadena de suministro debe facilitar y mantener el control sobre los flujos de mercancía. Es decir, debe propiciar suficiente flexibilidad para reaccionar a los rápidos cambios en la demanda del mercado. Factores tales como flexibilidad, rapidez y fiabilidad son de mucha importancia en el sistema de transporte (Ramírez, 2009).

Cabe resaltar que, en México, según la CANACAR (2017), dentro de los principales modos de transporte usados por las empresas, el que presenta un mayor índice de utilización en cuanto movimiento de mercancías, es el autotransporte de carga, esta modalidad representa el 55.7 % del flujo de productos y servicios, dejando en segunda posición al modo marítimo (31.3 %) seguido del ferroviario (12.9 %) y finalmente el modo aéreo (0.1 %).

Basado en el punto anterior, las siguientes secciones incorporan temas acerca de regionalización e infraestructura carretera, por ser tópicos inherentes en la pla-

nificación del transporte y que para el desarrollo actual y aplicación futura de esta investigación son aspectos clave para la creación de estrategias efectivas.

2.1.1 REGIONALIZACIÓN

Las regionalizaciones son clasificaciones espaciales especializadas que permiten simplificar y clasificar toda una complejidad territorial, que posee múltiples componentes e interrelaciones, a los más relevantes patrones socioeconómicos basados en concentración industrial, poblacional, riqueza económica, social y ambiental (CONABIO, 2010).

López (2018) indica que las regiones: (1) constituyen un marco holístico que integra las características significativas y fundamentales de los territorios; (2) ofrecen un marco para el análisis comparativo entre territorios que aporta un mejor conocimiento; (3) brindan un marco para la predicción y extrapolación y permiten acoger condiciones dinámicas; (4) son escalables, por tanto aplicables desde lo regional a lo local; (5) son un modelo robusto para el análisis gráfico de relaciones complejas.

Dentro de la literatura, Bassols-Batalla (1993); Cordero *et al.* (1977) consideran ocho regiones económicas en el país, ver figura 4.5, las cuales se definen en función de las actividades primarias, secundarias y terciarias, recursos naturales, el clima, demografía, la geomorfología y el territorio.



Figura 2.1: Regiones económicas de México

Fuente: México; Formación de Regiones Económicas (Bassols-Batalla, 1993).

Ríos y Obregón (2017) combinan la subdivisión del territorio en 8 regiones geoeconómicas con el sistema carretero federal del país con la intención de determinar niveles de accesibilidad intra-regional en la zona bajío, justificando su análisis con base a la extensión, cubrimiento y distribución que tiene el sistema carretero federal. El nivel de accesibilidad intra-regional depende de la demanda y de la distancia de los puntos de origen (en este caso almacenes) hacia cada región (clúster) donde se ubican los clientes. Por otra parte, Hanssens *et al.* (2003) mencionan que para llegar al mercado inter-regional la variable de accesibilidad se entiende como el potencial a la demanda que tienen los clientes hacia los almacenes, es decir, el mercado inter-regional sólo incluye la demanda potencial inducida por aquéllos considerados en dicha zona, bajo el supuesto de que tales municipios se comportan como una región económica homogénea.

Otro aspecto a considerar en la regionalización es la infraestructura y la cercanía al sistema de carreteras, la cual influencia la localización de los clústeres industriales, facilita el acceso al mercado y proveedores y permite una gestión eficiente

del transporte y distribución de productos acabados (Pardo y Carod, 2008; Biosca *et al.*, 2014). De acuerdo con la SCT (2017), en México la infraestructura vial está compuesta por cuatro tipos de carreteras (ver figura 2.2, donde solamente se puede hacer uso de estas con el tipo de vehículo reglamentado en la NOM-012-SCT-2-2017, que indica el tipo de configuración vehicular, pesos y dimensiones, lo que genera desafíos de movilización de mercancías procedentes de las empresas hacia el cliente, por tales razones, el conocimiento geográfico y carretero es un aspecto clave para el diseño de clústeres en la entrega de productos (Rendón y Morales, 2008). A continuación, se muestra el sistema carretero en México.

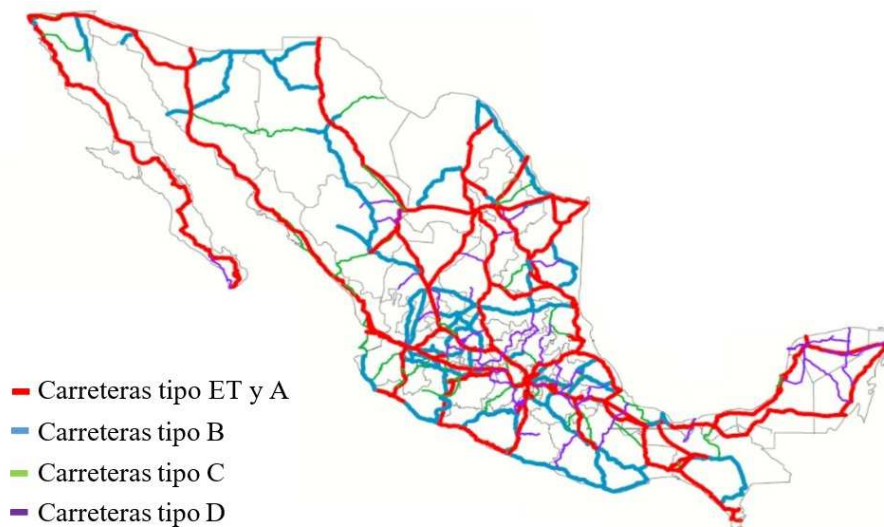


Figura 2.2: Clasificación de carreteras

Fuente: SCT (2017).

2.1.2 ACCESIBILIDAD A REGIONES

El concepto de accesibilidad, no posee una única y consensuada acepción, tampoco existe una medida de accesibilidad general ya que es entendida en términos geométricos (cercanía-lejanía), económicos, demanda, transporte, redes, entre otras perspectivas. Entonces la accesibilidad puede ser entendida como la facilidad con

que un servicio puede ser alcanzado desde una localización (Salado, 2004).

Algunas de las perspectivas para medir la accesibilidad son propuestas por Papa *et al.* (2015); Geurs y Van Wee (2004), por ejemplo:

- Medidas basadas en infraestructura: analizan el desempeño de la infraestructura del transporte como el nivel de congestión, tiempo y la velocidad promedio de viaje.
- Medidas basadas en ubicación: describen el nivel de accesibilidad a las actividades espacialmente distribuidas. Existen diferentes medidas de este tipo como las de distancia o las de gravedad que pondera las oportunidades de acuerdo a la impedancia, el tiempo de traslado o costos.
- Medidas basadas en utilidad: calculan el beneficio económico obtenido al acceder a ciertas actividades que están distribuidas espacialmente. La probabilidad que una persona elija cierta oportunidad va depender de la utilidad de la elección a la utilidad de todas las selecciones posibles.

Respecto a las medidas de accesibilidad existen variadas expresiones matemáticas, aplicables en SIG (sistema de posición geográfica), de las cuales Bosque (1994) señala:

- Medidas de accesibilidad basadas únicamente en la distancia.
- Medidas de accesibilidad basadas en la distancia y en el tamaño de la demanda en cada punto.
- Medidas de la accesibilidad basadas en el tamaño de la demanda dentro del alcance espacial del bien.

Por su parte, Salado (2004) indica tres medidas y establecimiento de áreas de influencia de equipamientos colectivos:

- Disponibilidad de equipamientos dentro de un alcance espacial dado.
- Accesibilidad considerando el tamaño de la oferta y los costos del viaje.
- Accesibilidad considerando la distancia y atributos variados de los centros de oferta.

En el trabajo propuesto por Ríos y Obregón (2017) se indica la función de accesibilidad al mercado, donde el potencial de demanda resulta del tamaño del mercado M_j dividido entre el costo de viaje C_{ij} . El efecto del costo se muestra inversamente proporcional al efecto del tamaño del mercado. Es decir, a mayor costo se incrementa el potencial de accesibilidad.

La estimación de la accesibilidad entre dos puntos puede realizarse de diferentes maneras, desde distancias directas hasta formulaciones complejas que consideren impedancia de traslado (Makri y Folkesson, 1999).

2.2 AGRUPAMIENTO (*Clustering*)

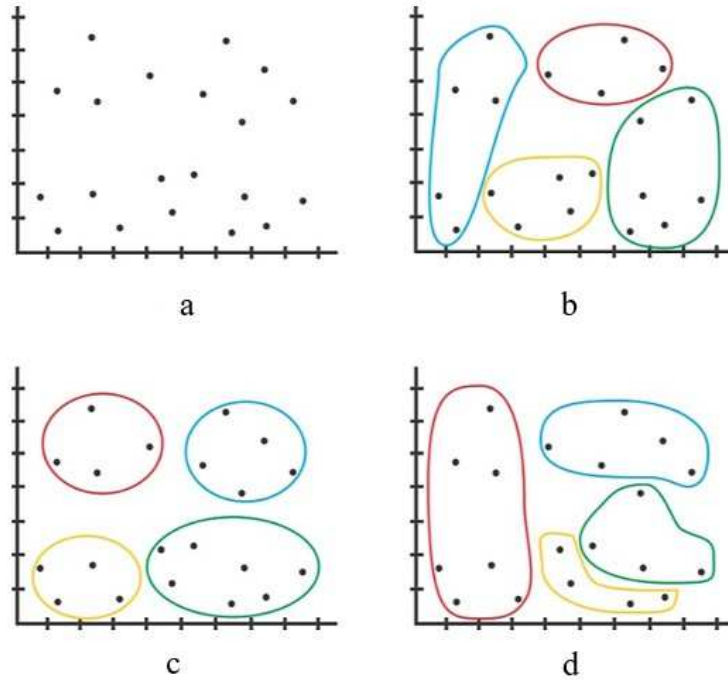
El análisis clúster es una de técnica de minería de datos, consiste en la división de los datos en grupos de objetos similares. Para medir la similaridad entre objetos se suelen utilizar diferentes formas de distancia: distancia Euclídea, de Manhattan, de Mahalanobis, etc. (Tian *et al.*, 2019). El problema de generar k clústeres de una población heterogénea de n elementos (clientes), es considerado como un problema NP-difícil cuando se consideran restricciones o características que incrementan su complejidad o cuando los puntos a ser agrupados se encuentren en un espacio euclídeo de dos dimensiones (Martínez y Cruz, 2011). Para estos casos, los métodos de solución sugeridos son los heurísticos, puesto que no se conoce un algoritmo determinístico en tiempo polinomial que los resuelva. Algunos de los métodos más utilizados son los conocidos como procesos no supervisados, enfocados al análisis de

conglomerados, los cuales son procedimientos que permiten encontrar similitudes y diferencias entre un conjunto de datos (Jain *et al.*, 1999).

De acuerdo con (Fung, 2001), una manera formal para describir el problema de agrupamiento es la siguiente: Dado $X \in \mathbb{R}^{m \times n}$ un conjunto de elementos representando un conjunto de m puntos x_i en \mathbb{R}^n , donde el objetivo es obtener la división de x en k grupos C^k , de modo que cada elemento perteneciente a un mismo grupo cuenta con características similares a los demás elementos de dicho grupo, mostrando diferencias sustanciales con elementos de grupos diferentes. Cada uno de los k grupos obtenidos durante la segmentación se conoce como grupo o agrupamiento. En la mayoría de los casos, los agrupamientos se realizan en base a la cercanía entre los elementos de una población, tratando de minimizar la distancia entre los elementos de un mismo agrupamiento.

De acuerdo con Araujo (2015), también podemos decir que el problema de agrupamiento se describe de la forma siguiente: dado un conjunto de puntos (elementos u objetos) $x = (x_1, x_2, \dots, x_n)$ de tamaño n , un clúster C_j es un conjunto de puntos, que basados en una medida de proximidad, son similares entre sí. El análisis aglomerativo es un proceso que permite dividir un conjunto en k grupos $c = (c_1, c_2, \dots, c_k)$ de datos distintos por medio de algún criterio de agrupamiento, como una función de costo, demanda, medida de accesibilidad o algún otro tipo de regla de asociación (Halkidi *et al.*, 2001).

Para ejemplificar el problema de agrupamiento, en la figura 2.3, se muestra un conjunto de elementos (datos) que se encuentran dispersos en un espacio euclídeo y para el cual se muestran tres de las posibles soluciones (agrupamientos) generados (a, b, c) para una instancia pequeña de 20 elementos (a) .

Figura 2.3: Agrupamiento (*Clustering*)

Fuente: Martínez y Cruz (2011).

Chen *et al.* (1996) señalan que el principio fundamental del agrupamiento es garantizar que los grupos sean lo más heterogéneos entre sí, pero que los elementos del grupo sean lo más homogéneos posibles, basados en un criterio de optimización. En otras palabras, lo que se busca es minimizar la distancia intra-cluster (cohesión) $\min d(x_1, x_2)$, y a la vez maximizar la distancia inter-clúster (separación) $\max d(c_1, c_2)$. La figura 2.4 expresa de manera gráfica lo expuesto anteriormente.

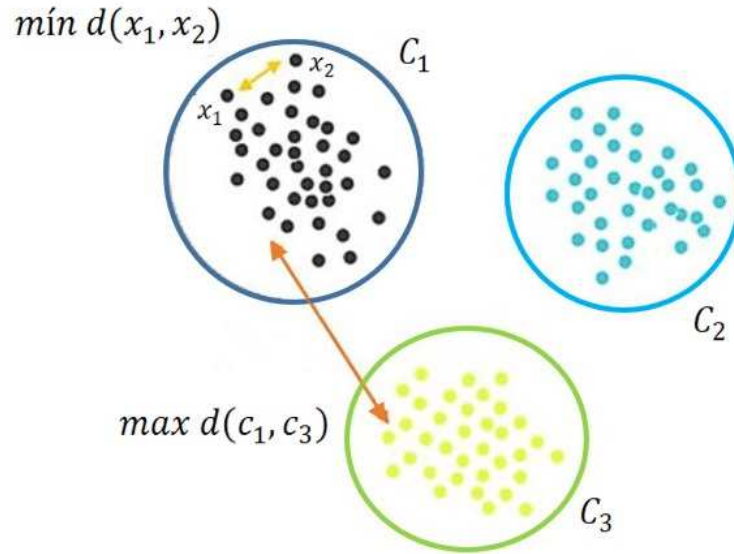


Figura 2.4: Características del clúster

Fuente: Vallejo Huanga (2016).

2.2.1 ALGORITMOS DE AGRUPAMIENTO

Son técnicas utilizadas en la minería de datos que intentan descubrir patrones en grandes volúmenes de datos mediante un proceso metodológico conocido como; descubrimiento de conocimiento implícito en base de datos, por sus siglas en inglés KDD. Estas técnicas también forman parte de algoritmos de aprendizaje no supervisado, es decir, buscan patrones en los datos sin tener una predicción específica como objetivo (no hay variable dependiente), por tanto, los datos solo tienen una entrada que serían las múltiples variables que los describen (Fayyad, 1997). Los algoritmos de agrupamiento pueden ser empleados para análisis estadísticos (análisis de grupos), reconocimiento de patrones, análisis predictivo y visualización para automáticamente extraer conocimiento (información), con el propósito de apoyar la toma de decisiones (Azevedo, 2019). Existen diversos algoritmos de agrupamiento propuestos y, su clasificación depende de la literatura que se revise. Pero de manera genérica los algoritmos de clasificación se pueden clasificar de acuerdo a Halkidi *et*

al. (2001) en:

- El tipo de datos de entrada para el algoritmo.
- El criterio de agrupamiento que define la similitud entre los objetos.
- La teoría y los conceptos fundamentales en los que se basan sus técnicas de análisis (ejemplo: teoría difusa, estadísticas, clasificación, reconocimiento de patrones).

De manera específica Pascual *et al.* (2007), clasifican los algoritmos de agrupamiento en: paramétricos y no paramétricos.

Entre los métodos de agrupamiento paramétricos se encuentran las mixturas finitas, éstas son una herramienta para modelar densidades de probabilidad de conjuntos de datos univariados y multivariados; modelan observaciones, las cuales se asume que han sido producidas por un conjunto de fuentes aleatorias alternativas e infieren los parámetros de estas fuentes para identificar qué fuente produjo cada observación, lo que lleva a un agrupamiento del conjunto de observaciones.

Los métodos de agrupamiento no paramétricos, no presuponen una forma concreta en el modelo a generar, resultando más flexibles y dando generalmente mejor resultado, a costa de requerir más datos para su ejecución y resultando más lentos que sus equivalentes paramétricos. Previamente Halkidi *et al.* (2001) había clasificado los algoritmos de agrupamiento no paramétricos en: jerárquicos, particionales y de densidad. En este trabajo se describen con mayor profundidad los dos primeros (jerárquico y particional) ya que son aquellos que realizan una división inicial de los datos en grupos apartir de la asignación arbitraria de un centroide, que itera repetidas ocasiones su posición entre lo objetos de un grupo a fin de minimizar la distancia entre casos.

2.2.2 ALGORITMOS DE JERÁRQUICOS

El objetivo de estos algoritmos es agrupar clústeres para formar uno nuevo o bien separar alguno ya existente, para dar origen a otros dos, de modo que, si el proceso se efectúa de manera sucesiva, este proceso minimiza o maximiza alguna medida de similitud (Yan, 2005). Una de las características principales en la agrupación jerárquica es que no requiere que se especifique el número de k clústeres.

El resultado del algoritmo es un árbol de clústeres, llamado dendograma (figura 2.5), que muestra cómo se forman y relacionan los clústeres. Cortando el dendograma a un nivel deseado (umbral), se obtiene una agrupación de los elementos en grupos disjuntos (Halkidi *et al.*, 2001). Los algoritmos de agrupamiento jerárquicos, de acuerdo al método que usan para formar los clústeres, pueden ser divididos en: divisivos o aglomerativos.

- Agrupamiento jerárquico divisivo - DHC (*Divisive Hierarchical Clustering*)

Se conocen como descendentes y construyen su jerarquía comenzando con un conglomerado que contiene todos los n objetos, y por medio de divisiones sucesivas, se van formando grupos cada vez más pequeños. En el estado final hay un total de k clústeres que contienen un solo objeto (Pascual *et al.*, 2007).

- Agrupamiento jerárquico aglomerativo - AHC (*Aglomerative Hierarchical Clustering*)

Son algoritmos ascendentes, constituyen el proceso inverso a los métodos divisivos. Comienzan el análisis con n grupos, formados por un solo objeto. A partir de estas unidades iniciales se van formando grupos, de forma ascendente, hasta que al final del proceso todos los objetos están agrupados en la misma agrupación. De las dos herramientas anteriores de acuerdo con Araujo (2015), comúnmente AHC es la más aplicada, ya que el proceso DHC es más complejo y más costoso en términos computacionales.

La figura 2.5, ejemplifica un dendrograma trazado para un conjunto de 40 puntos generados de manera aleatoria, bajo el método AHC y cortado a un umbral de 0.325 con la finalidad de obtener 7 grupos.

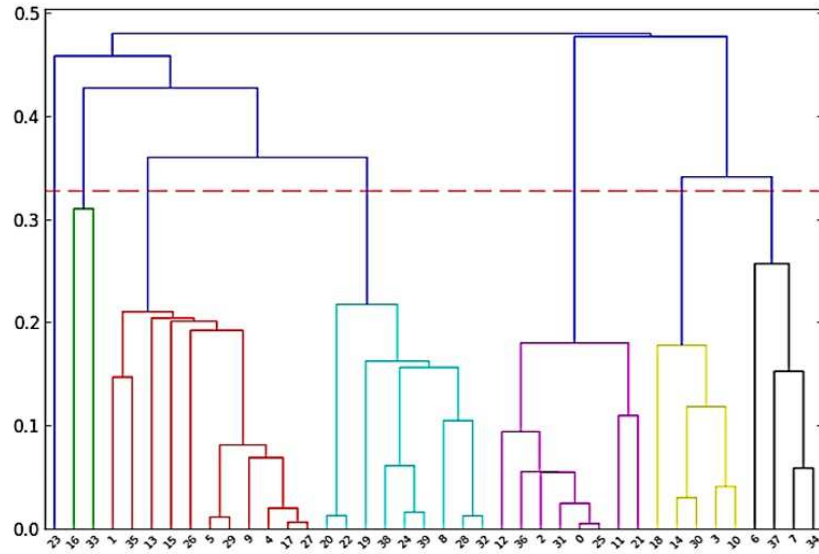


Figura 2.5: Dendrograma, AHC, $k=7$

Fuente:Vallejo Huanga (2016).

2.2.3 ALGORITMOS PARTICIONALES

En el agrupamiento particional, el objetivo es obtener una partición de los objetos en grupos de tal forma que todos los objetos pertenezcan a alguno de los k clústeres posibles y que por otra parte los clústeres sean disjuntos (Larrañaga y Lozano, 2001).

Estos algoritmos asumen un conocimiento a priori del número de clústeres en que debe ser dividido el conjunto de datos, llegan a una división en clases que optimiza un criterio predefinido o función objetivo. Entre los algoritmos más utilizados a nivel práctico podemos mencionar los siguientes:

PAM (*Partitioning Around Medoids*) usa k -medianas o aquellos casos repre-

sentativos en un conjunto de observaciones con el fin de identificar agrupamientos trabaja bien en bases de datos pequeñas, pero es lento en grandes (Guha *et al.*, 1998). El algoritmo CLARA (*Clustering Large Applications*) crea múltiples muestras de los datos y entonces aplica la técnica PAM a la muestra (Alfaro y Agustín, 2017). Por otra parte, el algoritmo con mayor aplicabilidad es el k -medias, crea agrupaciones a partir de un número ilimitado de casos que se basa en las distancias existentes entre cada uno de ellos en un conjunto de variables.

Una de las ventajas de los métodos particionales sobre los métodos jerárquicos se da en aplicaciones que implican el manejo de grandes conjuntos de datos donde la construcción de un dendograma es computacionalmente prohibitivo (Jain *et al.*, 1999).

Para el análisis de grupos, la técnica k -medias es la más observada en la literatura especializada (sección 2.4) por tal motivo se da a conocer a mayor detalle dicho algoritmo.

2.2.4 ALGORITMO k -MEDIAS

Según Kanungo *et al.* (2004), no se conocen algoritmos exactos de tiempo polinómico para este problema (agrupamiento), aunque hay distintas heurísticas que se utilizan en la práctica para tales fines. La herramienta k -medias, como ya se mencionó, es el algoritmo de clasificación de tipo particional más popular y aplicado en contextos reales, tiene como objetivo la partición de un conjunto de n observaciones en k grupos, en el que cada observación pertenece al grupo cuyo valor medio es más cercano, a fin de minimizar la suma de los cuadrados entre cada elemento y el centroide de su grupo (Tian *et al.*, 2019) basado en MacQueen *et al.* (1967). El procedimiento análisis de conglomerados de k -medias siempre utiliza para medir las distancias entre los casos, la distancia euclídea: la longitud de la recta que une ambos casos. MacQueen propuso y dio los pasos detallados del algoritmo k -medias.

Los pasos del algoritmo son los siguientes:

- Paso 1** Seleccione aleatoriamente k centros de clúster iniciales del conjunto de datos.
- Paso 2** Calcule la distancia de cada punto restante a cada centro de grupo de acuerdo con alguna función de distancia, y clasifique cada punto en la categoría del centro de grupo más cercano.
- Paso 3** Vuelva a calcular la media aritmética de cada grupo como un nuevo centro de grupo.
- Paso 4** Juzgando la convergencia o no, comparando el último y el segundo último centro del grupo, si no hay cambio, el grupo ha terminado, de lo contrario, continúe repitiendo los pasos 2 y 3.

Este tipo de algoritmo de aprendizaje no supervisado es útil para explorar, describir y resumir datos de una forma distinta. Utilizar este agrupamiento de datos nos puede servir para confirmar (o rechazar) algún tipo de clasificación previa. También nos puede ayudar a descubrir patrones y relaciones que desconocemos. Por ejemplo, podemos aplicar k -medias en: segmentación de clientes, agrupación de textos que hablan de temas similares, geoestadística, comunidades de redes sociales (Likas *et al.*, 2003).

2.2.5 COMPONENTES DE UNA TAREA DE AGRUPAMIENTO

De acuerdo Fayyad (1997), los pasos para el proceso de agrupamiento se pueden resumir en cinco, de los cuales los tres primeros son los que realizan el agrupamiento de los datos en clústers, mientras que los dos últimos se refieren a la utilización de la salida:

1. La selección de la técnica de clasificación y de las características, se refiere al proceso de identificar las características de los datos existentes (dimensión, tipo y volumen) e identificar el subconjunto más apropiado de características dentro del conjunto original de información para utilizarlo en el proceso de agrupamiento.
2. La proximidad de patrones se mide generalmente según una función de distancia definida para pares de patrones. Existen diferentes funciones de distancias que son utilizadas por distintos autores.
3. El agrupamiento de los objetos es la ejecución de los puntos anteriores.
4. La abstracción de datos es el proceso de extraer una representación simple y compacta de un conjunto de datos (de modo que la representación que se obtiene sea fácil de comprender e interpretable).
5. El análisis de validez de clústeres consiste en la evaluación de la salida obtenida por el algoritmo de clasificación. Este análisis utiliza a menudo un criterio específico.

Dice Villagra y Leguizamón (2007) que lo más común es calcular la diferencia o disimilitud entre dos patrones usando la medida de la distancia en un espacio de características. Existen varios metodos para definir la distancia entre objetos. La medida de distancia más popular es la distancia Euclídea, (ecuación 2.1), definida como:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2.1)$$

donde $i = (x_1, x_2, \dots, x_n)$ y $j = (y_1, y_2, \dots, y_n)$ son dos objetos de dimensión n . Otra métrica ampliamente utilizada es la distancia Manhattan (ecuación 2.2), definida por:

$$d(x, y) = (|x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|) \quad (2.2)$$

Tanto la distancia Euclídea como la distancia Manhattan satisfacen los siguientes requisitos matemáticos para la función de distancia:

$d(x, y) > 0$; si $x \neq y$. La distancia de x a y es mayor que cero, si y solo si x es diferente de y .

$d(x, y) = d(y, x)$. La distancia es una función simétrica.

$d(x, y) \leq d(x, h) + d(h, y)$. Se trata de una desigualdad triangular que afirma que ir directamente desde un punto x hasta un punto y nunca es más largo que pasando por un punto intermedio h .

Finalmente, la distancia Minkowski (ecuación 2.3) es una generalización de las distancias Manhattan y Euclídea. Se define por:

$$d(x, y) = (|x_1 - y_1|^q + |x_2 - y_2|^q + \dots + |x_n - y_n|^q)^{1/q} \quad (2.3)$$

donde q es un entero positivo, representa a la distancia Manhattan cuando $q = 1$ y a la Euclídea cuando $q = 2$. Un posible inconveniente que presenta la distancia de Minkowski es que las distancias de los atributos de mayor magnitud tienden a dominar al resto. Para solucionar esta desventaja se puede normalizar los valores de los atributos continuos, de forma que tomen valores dentro de un mismo rango.

2.2.6 VALIDACIÓN DE LOS GRUPOS

La evaluación de las agrupaciones es importante, ya que evalúa la calidad de las formaciones y es parte de cualquier análisis clúster. Una de las motivaciones es

que casi en todos los algoritmos aglomerativos se encontrarán grupos en el conjunto de datos, aun cuando este conjunto no contenga clústeres naturales en su estructura (Dalton *et al.*, 2009).

Las medidas de evaluación que se aplican para juzgar varios aspectos de la validez de los clústeres se clasifican generalmente en:

- No-supervisada: Mide la bondad de la estructura del agrupamiento sin referirse a información externa.
- Supervisada: mide el grado de coincidencia de la estructura de clúster descubierta por el algoritmo con alguna estructura externa conocida.

Con base en Dalton *et al.* (2009), existen diferentes aspectos a tener en cuenta para la validación de las clasificaciones obtenidas:

- Determinar cómo tienden a agruparse los datos del conjunto, es decir, distinguir si existe en los datos una estructura no aleatoria.
- Determinar el número correcto de clústeres.
- Evaluar cuán bien los resultados del análisis de un clúster se adecúan a los datos sin tener referencia de información externa.
- Comparar los resultados de un análisis de clúster con resultados externos conocidos.
- Comparar dos conjuntos de clústeres para determinar cuál es el mejor.

Los primeros tres aspectos no utilizan información externa, son técnicas no supervisadas, mientras que el cuarto aspecto requiere información externa. El último aspecto puede realizarse de ambas formas, supervisada o no-supervisada.

A continuación, se describirán las formas en la que el aprendizaje no supervisado evalúa los agrupamientos finales.

Las medidas para validar un clúster a menudo se dividen en dos clases: medida de cohesión de los clústeres, que determina lo cercanos que están los objetos dentro del clúster, y la medida de aislamiento, que determina la separación de un clúster con respecto a los demás (Dalton *et al.*, 2009).

Las medidas no-supervisadas a menudo se llaman índices internos debido a que usan solo información presente en el conjunto de datos. La cohesión de un clúster puede definirse como la suma de las proximidades con respecto al centroide del clúster (Villagra y Leguizamón, 2007). La cohesión está dada en la (ecuación 2.4) y la medida para la separación se muestra en la (ecuación 2.5).

Distancia dentro de grupos:

$$Intra_grupos = \frac{\sum_{i=1}^k \sum_{x,y \in C_i} \frac{d(x,y)}{n}}{k} \quad (2.4)$$

donde C_i representa el clúster i , k es la cantidad de clústers y n es la cantidad de elementos pertenecientes al clúster i .

Distancia entre grupos:

$$Inter_grupos = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d(c_i, c_k)}{\sum_{i=1}^{k-1} i} \quad (2.5)$$

donde c_i representa al elemento central (centroide) del clúster i , y c_j representa el centroide del clúster j y K es la cantidad de clústeres.

2.3 APLICACIONES DE ALGORITMOS DE AGRUPAMIENTO EN LOGÍSTICA

A lo largo de la revisión literaria se encontró que los metodos de clasificación se enfocan en brindar soluciones de inicio para facilitar la aplicación de otro método distinto para atacar un problema, siendo su alcance una de las razones principales por la que estos métodos juegan un papel importante en diferentes áreas de la investigación (Taillard, 2003). Son diversas las áreas del conocimiento que utilizan análisis de grupos, tales como; psicología y ciencias sociales, biología, estadística, planeación de la cadena de suministro, reconocimiento de patrones, procesamiento de imágenes (Sheu, 2007), y recientemente métodos clúster aplicados a minería de datos, donde se han aplicado a problemas de optimización combinatoria (Martínez y Cruz, 2011; Rivera *et al.*, 2013).

Los documentos referidos al agrupamiento de clientes dirigen esfuerzos a la creación e implementación de objetivos estratégicos y operativos, tal es el caso que presenta Ngai *et al.* (2009), con la finalidad de analizar y seleccionar los documentos científicos donde se han aplicado las técnicas de agrupamiento para la gestión estratégica de clientes basada en la identificación del cliente, atracción del cliente, retención del cliente y desarrollo del cliente. Por otra parte Sheu (2007) utiliza el método de agrupamiento difuso para responder eficientemente a la demanda y definición de estrategias operativas mediante el agrupamiento de clientes, determinación del servicio de entrega y optimización de las rutas de entrega considerando geografía estática, así mismo muestra que el rendimiento general de un sistema de distribución logística puede mejorarse en más del 20 % de acuerdo a los resultados obtenidos en un caso real.

En el trabajo mostrado por Hu y Sheu (2003), se presenta un nuevo enfoque para el desarrollo de estrategias avanzadas de distribución logística en respuesta a la creciente complejidad en los mercados, con el objeto de agrupar clientes antes

de ejecutar el enrutamiento de la flota en operaciones logísticas, la metodología se orienta hacia un estudio de caso y consiste en agrupar a los clientes en función de sus atributos de demanda. Así también la propuesta de Salazar *et al.* (2011), la cual trata sobre un modelo de programación bi-objetivo para construir bloques por zona donde los clientes tienen diferentes características de compra, su estudio está basado en un problema real de una empresa de distribución de bebidas en México, donde se quiere dividir el número total de bloques de una ciudad en grupos según a algunos criterios de planificación, con el objetivo de brindar apoyo a la toma de decisiones estratégicas y operativas en los procesos logísticos, además, la partición permite una gestión más eficiente de las ofertas de mercado, ya que reduce el número de clientes insatisfechos mediante la aplicación de ofertas especiales en cada bloque o territorio.

Tao (2002) proporciona un nuevo algoritmo de agrupación difusa no supervisada para agrupar los patrones de datos, sin información a priori sobre el número de agrupaciones, las suposiciones iniciales de las ubicaciones de los centros de clúster no son necesarias puesto que los múltiples centros se adaptan para representar la forma no esférica de los grupos. Mientras que Yang y Nataliani (2017), propone el denominado algoritmo c -medias difuso de aprendizaje robusto (RL-FCM), una herramienta de agrupamiento difuso que exhibe tres cualidades de robustez: (1) robusto a las inicializaciones (2) robusto a selección de parámetros, y (3) robusto al número de grupos (con un número desconocido de grupos), estas técnicas se han aplicado en procesamiento de imágenes, sistemas de ingeniería y segmentación de mercados.

En el trabajo presentado por Nallusamy *et al.* (2010), se aplica la clasificación de clientes de una empresa minorista en corea con el propósito de proponer estrategias a nivel gerencial de gestión de la relación con el cliente (CRM) segmentando clientes VIP e identificando patrones de compra utilizando los datos de transacción y las técnicas de minería de datos (k -medias, reglas de agrupamiento y asociación). Sin embargo, también se menciona que la mayoría de los estudios recopilados suelen centrarse en las grandes empresas.

Acorde al agrupamiento de clientes con fines estratégicos Pakyürek *et al.* (2018), personalizan los servicios y definen las estrategias de ventas a la cartera de clientes analizando sus datos existentes para determinar sus comportamientos de consumo y canales de adquisición de productos, preferencias específicas del producto, preferencia estacional etc. Mediante agrupaciones naturales basadas en la implementación de algoritmo k -medias y árboles de decisión, donde los grupos diseñados muestran cambios significativos, indicando que la solución es exitosa. El trabajo juega un papel importante en la creación de estrategias gerenciales y operativas para la cadena de suministro.

Recientemente Jintana y Mori (2019) presenta la aplicación de una técnica de análisis de datos para permitir a las empresas de mensajería servir mejor a los clientes y llevar a cabo la segmentación de los clientes de manera más efectiva a través de algoritmos de agrupamiento (k -medias) para la clasificación e identificación de clientes rentables con el propósito de implementar una gestión eficaz de la relación con el cliente que permita identificar segmentos de mercado y ayudar a los tomadores de decisiones a desarrollar planes estratégicos y operativos adecuados.

Para resolver el problema de agrupamiento se encontró un enfoque de solución híbrida jerárquica desarrollado por Díaz *et al.* (2010), que integra un procedimiento de agrupamiento heurístico en un marco de optimización. Se basa en la filosofía *cluster first - route second* para la cual sigue una metodología trifásica. Los grupos de nodos se definen primero, luego dichos grupos se asignan a los vehículos y, finalmente, el enrutamiento y la programación de cada recorrido individual en términos de los nodos originales. Los criterios de entrada utilizados para el procedimiento de agrupamiento son la demanda de los clientes, tipo de vehículo, distancias y tiempos de viaje entre nodos, tiempos de servicio y cargas de entrega. El objetivo del procedimiento es identificar un conjunto de grupos factibles que incluya todos los clientes.

En logística se utilizan también enfoques basados en agrupación para el pro-

blema de ubicación de instalaciones ejemplificado por Liao y Guo (2008), donde se transforma el problema de ubicación-asignación en un modelo de agrupación considerando capacidad suficiente e insuficiente de las instalaciones hacia los puntos de demanda. Dividen el problema en dos partes: (1) la asignación de demandas a las instalaciones que consideran limitaciones de capacidad y minimizan el costo; y, (2) la optimización iterativa de las ubicaciones de las instalaciones utilizando un método de agrupamiento k -medias adaptado que optimiza indirectamente la calidad de ubicación- asignación según las restricciones de capacidad de las instalaciones y los puntos de demanda. Con un enfoque determinista, el método basado en agrupación involucra menos parámetros que el método basado en GA (*Genetic Algorithm*) siendo más estable.

El algoritmo CCA (*Capacited Clustering Algorithm*), es una técnica de descomposición aplicada al tratamiento del Problema General de Distribución (GDP) analizado por Jintana y Mori (2019). El algoritmo propuesto consiste en descomponer el problema en subproblemas más pequeños, mediante la aplicación del método k -medias clásico, generando un número de agrupamientos definido, basados en la distancia, además de involucrar la generación de un centroide extra (gran centroide), el cual se encarga de ordenar y asignar nodos a grupos nuevos tomando en cuenta la capacidad del vehículo. Cabe mencionar que, durante esta fase, todos los centroides son recalculados, hasta que no existan elementos aislados, para este problema los métodos más utilizados debido a su eficacia y eficiencia para ciertos tamaños de instancias son el k -medias y sus variantes (Pakyürek *et al.*, 2018).

En el trabajo de Salazar *et al.* (2011), presentan un algoritmo de agrupamiento k -medias mejorado, aplicado al problema de ruteo vehicular con ventanas de tiempo (VRPTW), compuesto de 4 fases; la primera corresponde a la selección (inicialización) de los puntos de inicio o centroides, la segunda fase corresponde a la clasificación, asignación o distribución del conjunto de clientes en grupos, utilizando la regla del cliente o centroide más cercano, calcula la similaridad utilizando el error al cuadrado y el total del cuadrado del error con la finalidad de verificar la

variación existente dentro del agrupamiento, calcula la distancia entre los clientes utilizando la fórmula de la distancia euclidiana, con la finalidad de generar grupos que minimicen la distancia en su interior; la tercera fase corresponde a la fase de actualización, la cual se encarga de recalcular los centroides, para finalmente aplicar la fase correspondiente al criterio de paro del algoritmo.

A partir de un problema logístico real de la empresa relacionado con la distribución y comercialización de cigarrillos en Brasil y Cuba, Escalera y Baldoquín (2013) presentan un sistema de soporte a la decisión para el agrupamiento de clientes en zonas de distribución, por medio conjunto de métodos elaborados para agrupar clientes, debido a que no es posible utilizar métodos clásicos existentes de agrupamiento por requerimientos adicionales del problema, es decir considerando restricciones de capacidad en los clientes a agrupar. Para ello se optó seguir una metodología similar a la propuesta de Vallejo Huanga (2016); Ríos y Obregón (2017), los algoritmos son adaptaciones del conocido método *k*-medias con el uso de centroides, lo que le proporciona una clara interpretación física al problema abordado (agrupamiento en zonas geográficas), aportando el centro geométrico de cada clúster. Pacheco y Valencia (2005) proponen un método de 2 fases para la creación de clústeres, proporcionando de manera inteligente las semillas iniciales.

Los resultados obtenidos a cada problema dependerá del tipo de algoritmo de agrupamiento aplicado, así como de las características propias del problema, ya que cada algoritmo o método de agrupamiento cuenta con ciertas ventajas y desventajas, las cuales deben ser analizadas antes de aplicarlos a un problema en específico (Martínez y Cruz, 2011).

Las operaciones de recolección y entrega a gran escala se manejan mejor mediante la agrupación de clientes y la reducción de la red (Mesa y Ukkusuri, 2015). A su vez propone un marco sistemático para la agrupación de la demanda en la red de logística de carga. Su trabajo consiste en agrupar redes de transporte después elaborar una macro red de clientes utilizando las redes individuales, para finalmente

establecer conectividad entre la macro red con el objetivo ocupar la máxima capacidad del vehículo, por lo tanto, la topología (geografía y direccionalidad), las ganancias compartidas (volúmenes, costos y precios) son elementos clave para la agrupación de la demanda en las redes de logística de carga.

Para reducir los costos logísticos específicamente costos de embalaje Zhao *et al.* (2017) utilizó clasificación, se evaluaron 3 métodos, *k*-Medias, AHC y mapa de características de auto-organización con la finalidad de agrupar medidas de cajas corrugadas a medidas modulares de tal forma que en lugar de emplear 249 tipos de medidas se emplearan 92 mediante el método AHC debido a un mayor porcentaje de llenado promedio, concluyendo que se tiene una compensación entre la disminución del costo de compra y el costo de inventario del material y el aumento del costo de transporte.

Años después Ríos y Obregón (2017) analizan la accesibilidad viaria mexicana o red y la teoría de la localización industrial en razón a la demanda inter-regional e intra-regional de la región bajío justificando su análisis con base a la extensión, cubrimiento y distribución que tiene el Sistema Carretero Federal (SCF) en contraste con el sistema geo-natural donde zonas montañosas, planicies, cordilleras y penínsulas son determinantes hacia la decisión de haber trabajado sobre esta subdivisión de México (regiones geoeconómicas). Los resultados muestran que el potencial de accesibilidad debe considerarse directamente proporcional al tiempo de viaje e inversamente proporcional al tamaño del mercado (Yu *et al.*, 2016). Los métodos de clasificación se desempeñan de manera diferente según la dimensión, el tratamiento de los valores atípicos y la separabilidad de los grupos, tal como señalan Adolfsson *et al.* (2019), en esta autoría también se propone una metodología para evaluar la aptitud de los datos para ser agrupados: (1) el análisis de agrupamiento determina si los datos poseen una estructura de clúster inherente, (2) si los datos no poseen una estructura para ser particionados de manera significativa, entonces el agrupamiento puede no ser adecuado para el conjunto de datos, por el contrario, si los datos se pueden agrupar, se puede (3) seleccionar o desarrollar un algoritmo de agrupación, (4)

finalmente, la solución se valida aplicando medidas de calidad de agrupación como específica Villagra y Leguizamón (2007) y Ben-David y Ackerman (2009), que puede resultar en la selección de un algoritmo alternativo en caso de no haber encontrado un agrupamiento de buena calidad.

En la siguiente sección dejamos de hablar de clusterización para comenzar a hablar de pronósticos, con el propósito de integrar un marco teórico sólido, que permita dar soporte científico al caso de estudio y posterior aplicabilidad del mismo, mediante el uso de diferentes técnicas cuantitativas.

2.4 LA IMPORTANCIA DE LOS PRONÓSTICOS EN LA PLANEACIÓN

De acuerdo con Chiavenato y Sapiro (2017), la planeación puede ser considerada como una función administrativa para la organización, que permite la fijación de objetivos, el establecimiento de políticas, de procedimientos, y el desarrollo de programas para ejercer la acción planeada. Por tanto, la planificación permite llevar a cabo proyecciones y pronósticos con carácter prospectivo, lo cual posibilita anticipar comportamientos y reacciones de los mercados frente a decisiones y actitudes de las organizaciones (Montes *et al.*, 2016).

En el marco organizacional según Chiavenato y Sapiro (2017) existen tres horizontes de planeación:

- Planeación estratégica: corresponde a la planeación corporativa para toda la organización, con un enfoque estratégico es decir aquellas ideas trazadas al largo plazo con alto impacto económico. Involucra la gestión global, y es la que ha centrado nuestra atención en esta investigación.
- Planeación táctica: dirigida hacia la planeación funcional, es decir, para cada

departamento de la compañía. Toma en cuenta la gestión funcional. Realiza la conversión e interpretación de las decisiones estratégicas en planes concretos.

- Planeación operativa: orientada a las divisiones o cargos operativos de la organización, es decir, en nivel operacional.

En la planeación de las empresas los pronósticos son parte fundamental, ya que todos los departamentos de éstas elaborarán sus planes estratégicos y operativos, así como objetivos y presupuestos basados en ellos (Hanke y Wichern, 2006).

Los pronósticos pueden aplicarse a lo largo de la cadena de suministro, aunque es muy común su ejecución en mercadotecnia, producción, finanzas, recursos humanos y transporte. El pronóstico de transporte es la parte incondicional de la logística de salida, cálculos de capacidad, planes de viaje, disponibilidad y características del transporte. Se proyecta por varios horizontes de tiempo o intervalos de tiempo (Wilson y Keating, 1999).

El propósito básico de la planificación y gestión del transporte es hacer coincidir la oferta de transporte con la demanda de viaje, lo que representa la necesidad de planear dicho recurso. Es necesario comprender a fondo el patrón de viaje existente para identificar y analizar los problemas relacionados con el tráfico existente. También se necesitan datos detallados sobre el patrón de viaje actual y los volúmenes de tráfico para desarrollar modelos de predicción de viajes a determinado destino. La predicción de la demanda futura de viajes es una tarea esencial del proceso de planificación del transporte a largo plazo para determinar estrategias para satisfacer las necesidades futuras (Drličiak y Čelko, 2016).

En el sentido empresarial, un pronóstico es una herramienta que proporciona un estimado cuantitativo acerca de la probabilidad de eventos futuros que se elaboran en base en la información histórica o criterios de expertos (Pindyck y Rubinfeld, 2001). A diferencia de la anterior interpretación, Gutiérrez (2013), conceptualiza el pronóstico como aquella estimación del valor futuro de una variable mediante

la aplicación de métodos y procedimientos que contribuyen a reducir el margen de error, haciendo uso además del buen juicio y experiencia del responsable de realizar dicha estimación.

2.4.1 CLASIFICACIÓN DE LOS MÉTODOS DE PREDICCIÓN

Como antes se mencionó, para hacer la estimación futura de una variable, existen diferentes técnicas, estas se dividen en dos tipos básicos: métodos cualitativos y métodos cuantitativos. Se dice que el pronosticador que pueda lograr una combinación tanto de técnicas cualitativas como cuantitativas será más eficiente que aquél que no lo hace. En general, se recomienda evitar los extremos: quien basa sus predicciones sólo en juicio propio, o quien emplea únicamente métodos cuantitativos, el resultado obtenido significaría un pronóstico poco confiable o poco realista (Gutiérrez, 2013).

2.4.1.1 MÉTODOS CUALITATIVOS

Por lo general, en un método cualitativo se usa la opinión de expertos, quienes establecen de forma subjetiva un pronóstico de acuerdo a su juicio, experiencia y otros factores no numéricos que le dan sustento a sus predicciones. El uso de un método cualitativo se justifica cuando no hay disponibilidad de datos o si éstos son muy escasos, cuando los datos no son confiables, o bien, cuando existen datos, pero su obtención o acceso a ellos resulta demasiado difícil o costoso. También es empleado para enriquecer los resultados obtenidos a partir de un método cuantitativo (Shaub, 2019).

Algunos de los métodos cualitativos más conocidos se enlistan a continuación.

- Consenso de un panel

- Método Delphi
- Analogía histórica
- Investigación de mercados
- Pronóstico visionario.

A medida que el pronosticador adquiere experiencia, se van afinando sus estimaciones y muchas veces pueden llegar a ser, incluso, más precisas que las emitidas mediante algún método cuantitativo (Izar, 2007).

2.4.1.2 MÉTODOS CUANTITATIVOS

A diferencia de los métodos cualitativos, es deseable emplear un método cuantitativo cuando sí se dispone de información histórica confiable (datos). Estas técnicas requieren el estudio de dicha información para predecir el valor futuro de la variable de interés. Además, no es indispensable que la persona tenga experiencia, ya que en este caso no se necesita emitir un juicio basado en su intuición. Existen diversos métodos cuantitativos de pronóstico que se pueden clasificar en dos grupos, de acuerdo al tipo de datos con el que se cuente: métodos para pronosticar series de tiempo y modelos causales (Hanke y Wichern, 2006).

Los métodos de pronóstico para series de tiempo son sucesiones periódicas de datos históricos. El fundamento básico de estos métodos consiste en suponer que el comportamiento histórico de la variable seguirá teniendo el mismo patrón, por lo que se trata de proyectar hacia el futuro dicho comportamiento subyacente de la serie de tiempo. Para este tipo de métodos se clasifican en: métodos de suavización, métodos de descomposición y modelos autorregresivos (ARIMA).

Los métodos de pronóstico cuantitativos se mencionan a continuación.

Métodos de suavización

Dentro de esta clase de métodos encontramos:

- Series de tiempo estacionarias: promedio móvil simple; promedio móvil ponderado; suavización exponencial simple.
- Series de tiempo con tendencia: suavización exponencial lineal de dos parámetros (método de Holt).
- Series de tiempo estacionales: suavización exponencial lineal de tres parámetros (método de Winters).

Métodos por descomposición

- Descomposición multiplicativa
- Descomposición aditiva.

Modelos causales

Estos modelos se utilizan cuando se requiere identificar otras variables que de alguna manera estén relacionadas con la variable de interés. Se denominan causales debido a que por lo general se trata de variables que tienen una relación de causa-efecto, es decir, el comportamiento de una o más variables (la causa), determina en alguna medida el comportamiento de otra variable (el efecto). Este tipo de relaciones se pueden expresar mediante modelos matemáticos que estimarán el comportamiento de las variables de interés.

Dentro de esta clasificación, se encuentran principalmente los modelos de regresión y los econométricos.

Modelos de regresión

- Regresión lineal simple

- Regresión lineal múltiple.

Es importante saber identificar los diferentes patrones de comportamiento que puede presentar un conjunto de datos debido a que el método de pronóstico que se utilizará dependerá de ello, ver figura 2.6. Por tanto, las características que podemos encontrar en una serie de tiempo son:

- **Estacionaria.** Es aquella cuyo comportamiento general se observa dentro de una franja estacionaria u horizontal. En general no crece ni decrece a lo largo del tiempo.
- **Con tendencia.** Una serie de tiempo con tendencia presentará un comportamiento dentro de una franja ascendente o descendente.
- **Estacional.** Es la serie de tiempo con un patrón de comportamiento repetitivo en periodos iguales o menores a un año; es decir, el comportamiento anual se puede dividir en estaciones de igual magnitud. A esta división se le denomina estacionalidad que representa el número de periodos en que se divide cada año. Cabe mencionar que, si el comportamiento repetitivo a lo largo del tiempo se observa en periodos mayores a un año, se le denomina comportamiento cíclico. Una serie de tiempo puede ser estacionaria y estacional al mismo tiempo; o estacional con tendencia, pero no puede ser estacionaria y con tendencia a la vez.
- **Variaciones aleatorias.** Son saltos en los registros a causa de situaciones inusuales o por el mismo azar, son aspectos no controlables generados por imprevistos naturales, sociales o económicos. Son de corta duración y no se repiten.

Mencionadas las características que pueden estar presentes en una serie de tiempo, en la figura siguiente se analizan gráficamente cada una de ellas, teniendo como referencia la demanda actual (línea horizontal).

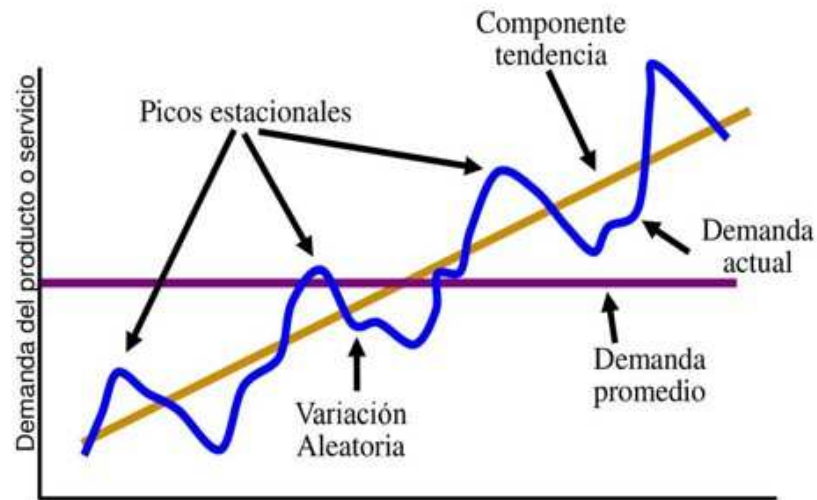


Figura 2.6: Patrones de comportamiento

Fuente:Gutiérrez (2013).

2.4.2 SELECCIÓN DEL MÉTODO DE PRONÓSTICO

Existen diversas técnicas para pronosticar, y hasta la fecha no hay manera de probar que alguno de ellos sea mejor que otro para este propósito. ? y Gutiérrez (2013) indican que cualquier modelo que sea seleccionado para la elaboración del pronóstico dependerá de diversos factores que hay que considerar:

1. Disponibilidad de datos: Permitirá la elección de un método cualitativo o uno cuantitativo.
2. Precisión deseada: Es importante definir qué nivel de exactitud se desea. La búsqueda de una mayor exactitud del pronóstico puede llevar al análisis de varios métodos y a otras consideraciones que pueden requerir principalmente mayor tiempo y recursos.
3. Uso que se le dará al pronóstico: Es necesario saber para qué se utilizará el pronóstico, por ejemplo, para establecer horizontes de planeación o comparar

un método contra otro.

4. Disponibilidad de recursos: Es de especial consideración tomar en cuenta los recursos disponibles. Es posible que se requiera gente preparada o de equipo especial para su elaboración.
5. Importancia del pasado para estimar el futuro: Es preciso considerar la relevancia del patrón de comportamiento histórico de la variable para identificar si puede tomarse en cuenta para el futuro.

2.4.3 MEDICIÓN DEL ERROR DE UN MÉTODO DE PRONÓSTICO

Existen diferentes mediciones de error que evalúan al pronóstico desarrollado. De acuerdo con Gutiérrez (2013); Perez *et al.* (2012); Torres-Rabello (2012), los más usuales son: (ecuación 2.7) el error medio (ME) y el error absoluto medio (MPE) (ecuación 2.9) que miden el sesgo en el pronóstico; o el error porcentual medio (MAE) (ecuación 2.8) y el error porcentual absoluto medio (MAPE) (ecuación 2.10) que dan información sobre el tamaño del error en términos porcentuales. Por otra parte el error cuadrático medio (MSE) (ecuación 2.11), es empleado para hacer comparaciones contra otros métodos utilizados, el cual castiga las diferencias entre un valor real y un estimado elevándolas al cuadrado. Sin importar los métodos de pronóstico desarrollados, la labor del pronosticador es encontrar aquel que arroje el error mínimo para garantizar una mayor precisión en la estimación, y por consecuencia una mejor planeación y toma de decisiones (Torres, 2011).

Se le llama error a la variación que existe entre el valor real de la variable y su valor pronosticado (ecuación 2.6). El error se define entonces como:

$$e_t = Y_t - F_t \quad (2.6)$$

Las siguientes mediciones de error se aplican a los pronósticos de series de

tiempo, aunque su comprensión será útil también cuando se trate de pronósticos mediante modelos causales.

Error medio (ME):

$$ME = \frac{\sum_{t=1}^n (Y_t - F_t)}{n} \quad (2.7)$$

Error absoluto medio (MAE):

$$MAE = \frac{\sum_{t=1}^n |Y_t - F_t|}{n} \quad (2.8)$$

Error porcentual medio (MPE):

$$MPE = \frac{\sum_{t=1}^n \left(\frac{Y_t - F_t}{Y_t} \right)}{n} \quad (2.9)$$

Error porcentual absoluto medio (MAPE):

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - F_t}{Y_t} \right|}{n} \quad (2.10)$$

Error cuadrático medio (MSE):

$$MSE = \frac{\sum_{t=1}^n (Y_t - F_t)^2}{n} \quad (2.11)$$

donde:

e_t = error del pronóstico en el periodo t

Y_t = valor real de la variable en el periodo t

F_t = valor pronosticado de la variable en el periodo t

n = número de periodos.

La evaluación de una estimación mediante el cálculo del error, indicará que tan validos son los resultados obtenidos. Un pronóstico debe sobrepasar el 70 % de exactitud para que sea confiable y se evite construir estrategias inadecuadas que puedan causar aumento en costos, disminución en el nivel de servicio y pérdida de imagen a la empresa (Lu y Kao, 2016).

Como característica inherente de un pronóstico, el juicio y la intuición del pronosticador influirán en el método a elegir y, por consecuencia, en el resultado de la estimación. En resumen, no existe un método que sea mejor que otro; simplemente dependerá de cada situación. Podría decirse que cada caso tiene su propio “mejor método” y este será el que elija el pronosticador, una vez que haya puesto a consideración los factores antes mencionados.

Los métodos de agrupamiento como algoritmo de inicialización para la aplicación de pronósticos puede evitar problemas tales como: disponibilidad excesiva de unidades, vehiculos faltantes, altos costos por renta motriz, costos adicionales por renta inmediata, eventualmente incremento del nivel de servicio aproximadamente en 20 %, mejorando así la eficacia de la gestión de la cadena de suministro y generando ahorros sustanciales en la subcontratación del servicio de distribución a partir de pronósticos precisos y sólidos (Sheu, 2007).

La idea fundamental del modelo de pronóstico basado en clasificación es utilizar un algoritmo de agrupamiento para dividir el total de clientes de una empresa en múltiples grupos disjuntos y construir un modelo de pronóstico para cada grupo basado en el análisis de la información subyacente de cada cliente (Lu y Kao, 2016). Los datos se asignan a un grupo por su similitud, y el modelo de pronóstico de un grupo particular se utiliza para obtener resultados de pronóstico para ese grupo.

Debido a que los datos en el mismo grupo tienen patrones de datos similares, el modelo de pronóstico basado en agrupamiento puede producir una mejor precisión de pronóstico que el modelo de pronóstico construido sobre un conjunto de datos completo.

Con base a esta información el presente trabajo está dirigido hacia la planificación efectiva del transporte a través de herramientas de clasificación y de predicción, cuyos procedimientos son de gran interés en estadística.

2.4.4 CONCLUSIÓN DEL CAPÍTULO

Este capítulo fue diseñado para generar una base teórica que brinde el soporte académico apropiado para el desarrollo de capítulos posteriores, además se manejaron los conceptos y se habló de aplicaciones en el campo logístico, con la finalidad de mostrar al lector las diferentes posibilidades de atacar un problema similar al expuesto en este trabajo, y de guiar a nuestra empresa caso de estudio para resolver complicaciones actuales y futuras. Del mismo modo, de acuerdo a lo establecido en la literatura consultada, realizar planes de transporte y distribución sin fundamento cuantitativo y análisis previo, puede ocasionar inconsistencias en los días de inventario, monetización retardada, ventas perdidas, sobrecostos e insatisfacción del cliente.

En el capítulo siguiente, se retomarán algunos de los conceptos anteriores a fin de diseñar el procedimiento que nos permitirá alcanzar el objetivo planteado en el primer capítulo.

CAPÍTULO 3

METODOLOGÍA

En este capítulo se expone la metodología a seguir para lograr planeaciones estratégicas y financieras confiables, utilizando herramientas cuantitativas de agrupamiento y predicción, que permitan anticipar la demanda de transporte de carga a clústeres de clientes con características similares. Se describe el proceso de resolución para la problemática a partir de la revisión de literatura mostrada en el capítulo 2. También es descrito el procedimiento que servirá de guía para alcanzar el objetivo perseguido por nuestro caso de estudio, de tal manera que en el capítulo 5 se verifique la viabilidad de la metodología por medio del análisis de resultados arrojados en el capítulo 4 implementación.

3.1 DESCRIPCIÓN GENERAL

La metodología sugerida y esquematizada en la figura 3.1, está basada en la bibliografía especializada y ha sido adaptada para nuestro caso de estudio, por lo tanto, con la finalidad de diseñar un procedimiento completamente pragmático, actividades como la identificación del problema y áreas de oportunidad así como la revisión de la literatura, son fases previas no mostradas en la esquematización, pero que representan el sustento científico y la dirección de la investigación, sin embargo para considerar su aplicación en los distintos procesos de la cadena de suministro,

son fases que no deben pasarse por alto. La figura metodológica principal comprende 5 fases, la primer fase consiste en seleccionar las variables que serán empleadas para la clusterización, la segunda fase se centra en la elección del algoritmo de agrupamiento de acuerdo al tipo y volumen de datos a procesar, la tercer fase trata sobre la validación de los grupos sugeridos por el algoritmo de agrupamiento, la cuarta fase se enfoca en el diseño de clústeres, finalmente la quinta fase consiste en la construcción y aplicación de modelos de predicción que permitan cuantificar la demanda futura de transporte a nivel clúster. A continuación, se muestra el esquema secuencial de lo antes mencionado.

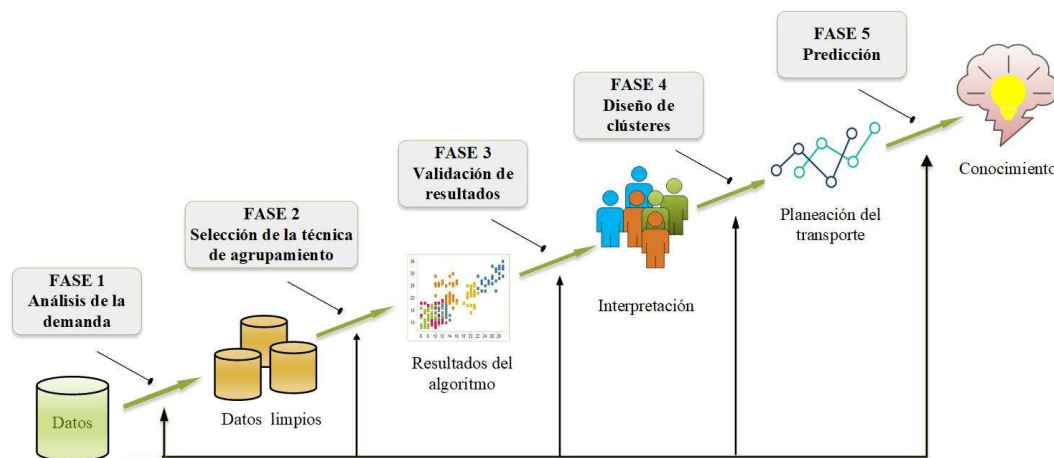


Figura 3.1: Esquema metodológico

Fuente: Elaboración propia basada en Vallejo Huanga (2016); Halkidi *et al.* (2001); Fayyad (1997).

3.2 FASES PRELIMINARES

Identificación del problema y áreas de oportunidad: Debemos identificar los factores que denotan la presencia de una situación de desventaja en la cadena de suministro, en este caso el transporte de carga, esto con la finalidad de analizar áreas en donde es posible mejorar los procesos logísticos, tomando en cuenta la alineación

con los intereses de la empresa.

Revisión de literatura: Una vez identificado el problema que se abordará, se realiza una revisión de la literatura fundamentada en conceptos, definiciones y problemáticas similares encontradas en trabajos de investigación, libros, reportes técnicos y estadísticas de información, a fin de mostrar un panorama cronológico sobre los estudios realizados y sobre los aspectos que continúan desconocidos. En la figura 3.2 se ilustra el proceso utilizado para la revisión bibliográfica, el cual consistió en recabar información mediante criterios específicos de búsqueda y filtrado (palabras clave, año, autor e idioma) constituyendo el método de revisión sistemática de literatura (García, 2017).

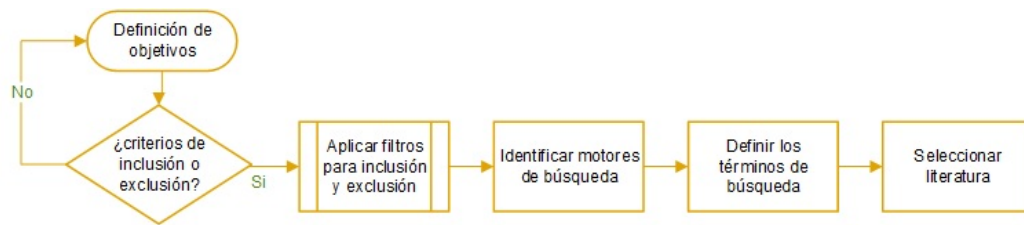


Figura 3.2: Revisión sistemática de literatura

Fuente: Elaboración propia basada en García (2017).

3.3 FASE 1: ANÁLISIS DE LA DEMANDA

En esta fase se examina el comportamiento de la demanda de transporte para cada cliente a través del tiempo, para realizar este análisis es imprescindible contar un gestor de base de datos capaz de procesar un volumen grande de información como lo es *SQL server*, o herramientas de procesamiento *POWER BI*, que encontramos en la interfaz de Microsoft Excel. El análisis de la demanda implica la recolección de datos y la limpieza de los mismos, a fin de facilitar la extracción y manipulación de las variables que alimentarán al algoritmo de agrupamiento y de predicción. Una vez listos los datos, se deben seleccionar las variables que participarán en fases siguientes,

las características elegidas no siguen ningún criterio de selección, sino se emplearán las variables disponibles en la base datos y las que mejor relacionadas estén con el objetivo de la empresa o del usuario. En esta investigación son consideradas las variables recomendadas por la literatura especializada (capítulo 4), las cuales en su mayoría están presentes en la información proporcionada por nuestro caso de estudio. Es importante resaltar que, la mayoría de los algoritmos de agrupamiento funcionan cuando menos con una variable, sea de tipo numérica o categórica. A medida que mayor sea el número de variables, los clústeres resultantes serán más afines, sin embargo, para generar un escenario gráfico se tendrán que definir aquellas variables que describan la mayor parte del problema de agrupamiento a través del análisis de componentes principales, por sus siglas en inglés PCA (*principal component analysis*), que permita la visualización en un plano de dos dimensiones.

3.4 FASE 2: SELECCIÓN DE LA TÉCNICA DE AGRUPAMIENTO

Esta fase consiste en determinar la técnica de agrupamiento apropiada y en la ejecución del algoritmo. Por lo tanto, es necesario identificar el tipo de agrupamiento que se busca realizar (jerárquico, particional o densidad), considerando propiedades de escalabilidad, dimensionalidad y robustez, las cuales son propiedades que deberá mantener toda base de datos con la aspiración de clasificar un conjunto de observaciones mediante un análisis clúster de tipo particional. También es importante analizar lo encontrado en la bibliografía especializada, ya que, el método más utilizado para resolver el problema de agrupamiento de clientes, será el criterio de selección para nuestra investigación, adoptando así la técnica comúnmente desarrollada por los expertos y, que en cuestiones de factibilidad resulte económicamente y computacionalmente posible, de manera que sea también un algoritmo que proporcione resultados efectivos, los algoritmos con mayor incidencia aplicada y documentada por los investigadores son: k -medias, k -medianas, CLARA (*clustering large applica-*

tions) y variantes de k -medias. Normalmente los algoritmos particionales recurren a una medida de distancia que permita establecer la dispersión o cohesión de un clúster, esta medida está dada de acuerdo a la técnica a ejecutar, por lo general la literatura recomienda emplear distancia Euclidiana, Manhattan o Mahalanobis, así también, por su practicidad, interfaz gráfica, efectividad y facilidad de interpretación, es recomendable emplear el paquete informático R studio para la ejecución del algoritmo de clusterización.

3.5 FASE 3: VALIDACIÓN DE RESULTADOS

Una vez llegado a esta fase de la metodología, se comprueba la homogeneidad y heterogeneidad de los grupos obtenidos de la fase anterior, aplicando alguno de los índices de evaluación de grupos descritos en el capítulo 2 (índice de la silueta, índice de Dunn o el índice de Davies), los cuales son frecuentemente recurridos ante problemas de agrupamiento. Es decir, los resultados de la aplicación de un algoritmo de agrupamiento deben ser verificados por medio de criterios y técnicas adecuadas, puesto que, al tratarse de técnicas no supervisadas, los grupos finales son validados con métodos independientes a las herramientas cuantitativas de agrupamiento.

3.6 FASE 4: DISEÑO DE CLÚSTERES

Esta fase corresponde a la creación de los clústeres finales, se basa en el análisis de resultados del algoritmo y al análisis interpretativo que involucra el efecto que tuvieron las variables seleccionadas sobre las observaciones (clientes) en la formación de regiones, con la finalidad de establecer relaciones analíticas y graficas que expliquen la causa de la formación, basándonos en la distancia de cada partición a su respectivo centroide. Debemos tener en cuenta que para el diseño de clústeres es necesario haber normalizado previamente los datos para evitar datos atípicos.

3.7 FASE 5: PREDICCIÓN

Es la fase final de la metodología propuesta y, consiste en la construcción de modelos predictivos para cada uno de los clústeres obtenidos en la fase 4. Por la naturaleza nuestros datos, los métodos de pronóstico susceptibles a utilizar pertenecen a los de tipo cuantitativo, ver figura 3.3, por lo tanto, las series de tiempo serán las estructuras básicas que nos permitirán cuantificar la demanda de transporte para clientes afines. En esta fase buscamos medir el desempeño de distintos métodos de predicción que sean compatibles con las series de tiempo presentes en cada clúster, siendo la variable de interés la demanda de transporte por clase de vehículo de carga. En el capítulo de resultados, se especifican los métodos que serán evaluados y donde el mejor modelo predictivo será aquel que minimice el error entre valores reales y ajustados.

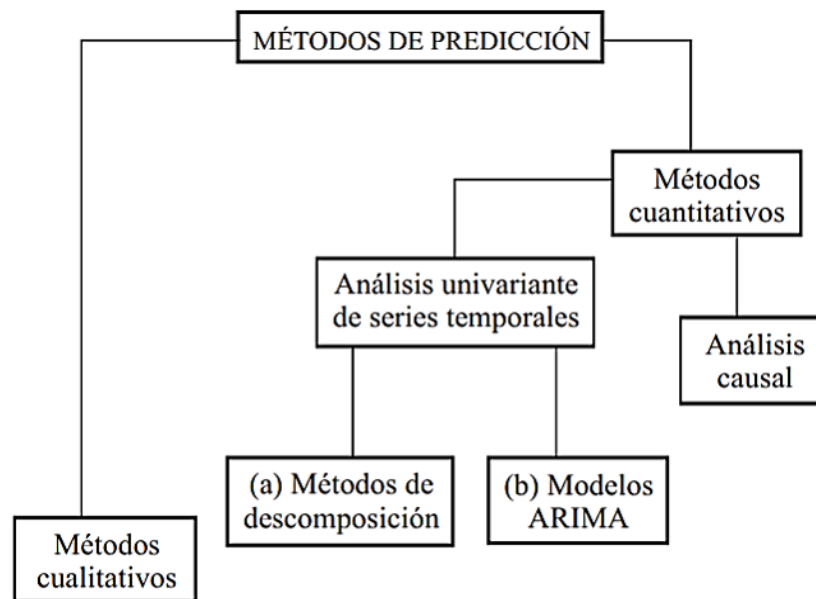


Figura 3.3: Clasificación de los modelos de predicción

Fuente: Guerrero *et al.* (2006)

La selección y aplicación de los modelos predictivos dependerá del comportamiento tendencial, estacional, cíclico y aleatorio que presente la serie de tiempo

correspondiente a cada clúster de clientes, dichas características son mencionadas en el capítulo 2.

3.8 CONCLUSIONES DEL CAPÍTULO

Este capítulo fue diseñado con la finalidad de proponer un procedimiento general que sea adaptable a casos de estudio que presenten problemas logísticos comunes a los abordados en esta tesis. Para encaminar la investigación a la resolución exitosa del caso de estudio, la metodología propuesta destaca cinco fases, partiendo de la exploración y conocimiento de áreas de oportunidad, tratamiento de datos, diseño de clústeres hasta la proyección de la demanda de transporte; fases que en conjunto establecen las bases para generar planes tácticos y financieros precisos. Una vez que se revisa la propuesta metodológica, el siguiente capítulo se enfoca en la aplicación con datos del entorno real, para observar el desempeño que tienen las técnicas de agrupamiento y las predictivas de forma independiente y en conjunto, a fin de confirmar la efectividad de estas herramientas matemáticas y a fin de establecer conclusiones que ayuden a nuestra empresa caso de estudio y posiblemente a otras organizaciones a solucionar una problemática en común. considerando a su vez los objetivos de la empresa y propiamente del objetivo de esta tesis.

CAPÍTULO 4

IMPLEMENTACIÓN Y ANÁLISIS DE RESULTADOS

De acuerdo a la metodología desarrollada en el capítulo 3, se presentan la implementación y los resultados obtenidos de cada fase del procedimiento propuesto, se describen las herramientas computacionales utilizadas para el agrupamiento de clientes y predicción de la demanda de transporte, también son descritos los experimentos realizados así como la información que implica su aplicación.

4.1 FASES PRELIMINARES

El desarrollo del proyecto comienza con las fases preliminares, las cuales consisten en la identificación de áreas de oportunidad y la revisión sistemática de la literatura.

4.1.1 IDENTIFICACIÓN DE ÁREAS DE OPORTUNIDAD Y DEFINICIÓN DEL PROBLEMA

Analizando los componentes del costo logístico presentes en actividades tales como: inventarios, almacenaje, compras, logística inversa y proveduría (recordar sección 1), podemos determinar a simple vista que cualquier actividad representa un área de oportunidad. Sin embargo, por significar el costo más elevado, la prioridad se enfoca en el dirigir esfuerzos en actividades de transporte y distribución. En función de nuestro caso práctico también se identificaron y analizaron áreas de oportunidad específicas relacionadas con los procesos logísticos que impactan no solo al transporte sino al resto de actividades logísticas antes mencionadas, las áreas de oportunidad se enlistan a continuación.

Control del abánico de clientes: Para la definición de planes tácticos donde se toman las decisiones de transporte y distribución, es indispensable conocer las necesidades del cliente y sus características (volúmenes, patrones, inversiones y condiciones de compra del cliente), a fin de controlar las operaciones que se verán involucradas en la cadena de suministro. En la empresa bajo estudio, se ha observado que el control del abanico de clientes dificulta el proceso de planeación de transporte, lo que resta visibilidad a la cadena de suministro, principalmente en operaciones relacionadas al análisis de la demanda, que, enfocado al proceso de transporte encontramos dificultad en la estimación de cajas y determinación de la clase de vehículo a emplear y, por lo tanto, aumento en costos de transporte y disminución en el nivel de servicio.

Regionalización: Actualmente la empresa caso de estudio establece el control de las operaciones acordes a la entrega de productos terminados al cliente basado en la división de regiones geográficamente diseñadas. Por lo que, la segmentación de las regiones queda justificada en criterios de ubicación, dando la posibilidad de basar estratégicamente la formación de regiones en el análisis de grupos utilizando criterios técnicos (por ejemplo, el costo y la demanda), a fin de facilitar la toma de decisiones en procesos productivos, mercadológicos y logísticos, enfocados hacia un conjunto

de clientes teóricamente estructurados.

Agrupamiento de clientes: Agrupar la demanda es un área de oportunidad latente que, para nuestro caso de estudio representa un desafío para la identificación de clientes críticos, atracción de nuevos clientes y retención de los actuales. Agrupar un alto número de clientes (que en nuestro caso de estudio está entre 1000 y 2000 clientes mayoristas) permitirá monitorear, clasificar y medir el comportamiento de los consumidores y, en consecuencia, profundizar en el conocimiento de los mismos al punto que se les pueda atender en forma segmentada e incluso personalizada.

Alineación de procesos: En la descripción general de problema (ver capítulo 1, sección 1. 1) se menciona la desalineación interna existente entre los departamentos responsables de la gestión del transporte y de la gestión financiera, por tal motivo, un área de oportunidad es la alineación de procesos. En cadena de suministro la integración interna se alcanza cuando los participantes de las áreas funcionales comparten su información para lograr un objetivo común: cumplir con la demanda bajo las condiciones de cantidad y lugar, de la manera más eficiente posible, utilizando de forma efectiva y compartida los recursos disponibles.

Planeación de la demanda de transporte: Son diversos los nodos que solicitan productos terminados, sin embargo, existen clientes que demandan un servicio personalizado en cuanto al tipo de vehículo, carga no fraccionada y tiempo de servicio. Cuando se visita una zona es normal encontrar clientes exigentes y clientes flexibles, por lo tanto, encontrar un balance entre este tipo de público dificulta la planeación de las unidades que serán rentadas para movilizar las órdenes de compra. Por otra parte, la anticipación con la que deberá realizarse el plan de transporte requiere semanas de preparación, por lo que hace de este punto un área de oportunidad presente en los procesos mensuales de planeación (S&OP) y que es susceptible a la aplicación de modelos predictivos para aproximar el número y la clase de vehículos a emplear.

Definición de métricos de desempeño: Para buscar mejorar un sistema es necesario conocer su desempeño. Se identificó esta área de oportunidad porque a pesar de

contar con métricos efectivos, es posible definir medidas de desempeño adicionales que permitan controlar otros aspectos de la operación en el proceso de transporte y nos ayuden a elaborar planes de acción concretos. Algunos métricos muy importantes a incluir son: el nivel de precisión de las predicciones, el nivel de error de las estimaciones y el nivel de utilización de vehículos por cliente o por clúster de clientes, con el fin de gestionar planes confiables de transporte y distribución para que no se vea afectada la estructura de costos o nivel de servicio.

Adopción tecnológica: La implementación de diferentes tecnologías acordes a los objetivos y actividades de la empresa requiere de la adopción de herramientas de computo que aumenten la eficiencia en el análisis de grandes volúmenes de datos y permitan obtener resultados en tiempo real optimizando procesos específicos y generando información oportuna.

4.1.2 REVISIÓN SISTEMÁTICA DE LA LITERATURA (SRL)

Esta fase constituye la investigación sobre las áreas de oportunidad a explorar. Por razones de prioridad en costo y operación, se decidió incursionar en dos de las áreas de oportunidad detectadas en la fase anterior. La primer área de oportunidad va encaminada a controlar el abanico de clientes y la segunda a mejorar el proceso de planeación del transporte.

En la tabla 4.1 se muestran las fuentes de consulta donde se recabó la teoría especializada para solventar ambas áreas de oportunidad. Siguiendo el proceso de la revisión sistemática de la literatura mencionado en el capítulo 3, la búsqueda se abordó de la manera siguiente:

Tabla 4.1: Portales literarios

Fuente de consulta	Frecuencia
Google scholar	32 %
Springer	20 %
ScienceDirect	13 %
Emerald	7 %
Scielo	10 %
Otros	18 %

Fuente: Elaboración propia

La búsqueda se realizó bajo los siguientes criterios de inclusión: algoritmos de agrupamiento, aplicaciones de las herramientas de agrupamiento, utilidad de los algoritmos, proceso de planeación, transporte en la cadena de suministro, pronósticos, medidas de desempeño. Así también, se utilizaron los siguientes términos de búsqueda: documentos del año 1980 al 2019; idiomas inglés y español; artículos publicados en revistas indexadas e informes técnicos de universidad con nivel igual o mayor a maestría.

Definidos los criterios para la *SRL*, se realizó la búsqueda y la selección de la bibliografía especializada que servirá como fundamento y guía para el desarrollo de las fases de la metodología.

La revisión nos permitió encontrar diferentes usos y aplicaciones de los algoritmos de agrupamiento, en la figura 4.1, se muestran los enfoques prácticos que los autores consultados abordaron en sus investigaciones.

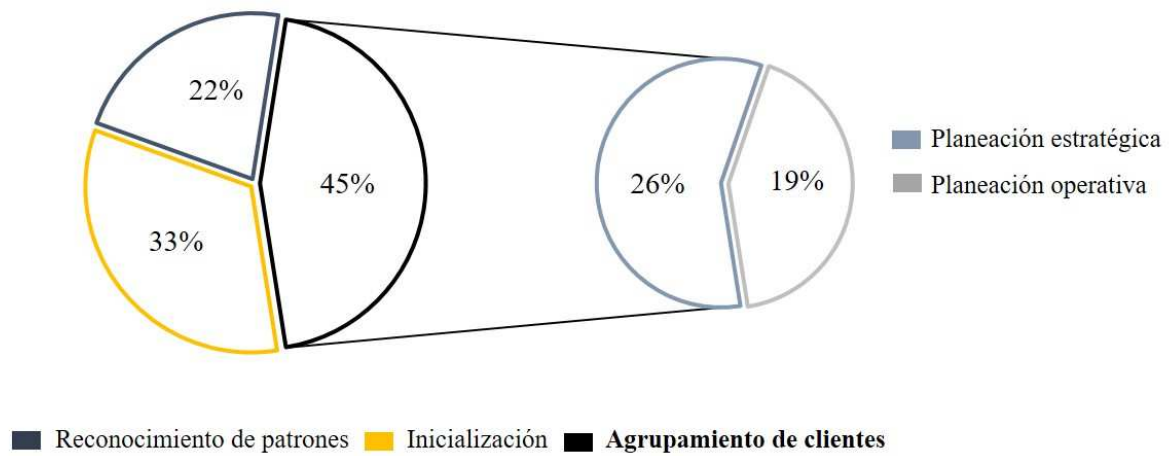


Figura 4.1: Enfoques de agrupamiento

Fuente: Elaboración propia basada en la literatura

Como parte de los resultados de la revisión de literatura realizada encontramos que el 22 % de los documentos consultados se enfocan en desarrollar agrupamientos para realizar reconocimiento de patrones, la cual es una técnica de aprendizaje no supervisado que busca encontrar propiedades específicas de un conjunto de observaciones físicas o abstractas a fin de determinar si observaciones posteriores poseen las características necesarias para considerarse dentro del conjunto, únicamente basado en el análisis computacional de una imagen, una textura, un comportamiento que clasifique los resultados en distintas categorías o agrupaciones. El 33 % de la bibliografía especializada revisada enfoca su trabajo en la inicialización de problemas de combinatoria, donde agrupar primero un conjunto de elementos resta complejidad al problema original de optimización, tal como sucede con problemas de enrutamiento de vehículos y localización de instalaciones. La filosofía agrupar primero-rutear después, mejora la solución final de 5 % hasta 20 %.

Por último, el 45 % de los autores consultados emplean las técnicas de agrupamiento para agrupar clientes con la finalidad clasificar el mercado para establecer

estrategias concretas que impacten a un segmento de mercado en particular de acuerdo a las características que predominen en la agrupación. Se encontró que a través del análisis clúster enfocado al agrupamiento de clientes el 26 % de los autores se inclinan a la planeación estratégica, es decir, a la planeación de actividades relacionadas con la alta dirección encargada de tomar decisiones acordes a la ubicación y apertura de nuevas plantas y centros de distribución, selección del modo de transporte, control de inventarios, diseño del procesamiento de pedidos, establecimiento de estándares, desarrollo de proveedores. Por otra parte, el 19 % de documentos están orientados a la planeación operativa, es decir a las actividades relacionadas con la toma de decisiones a nivel operacional y jefaturas por ejemplo establecer política óptima de inventarios, despacho, procesamiento de pedidos, liberación de pedidos y facturación.

Según los resultados de la revisión bibliográfica, la mayoría de los trabajos se enfocan al agrupamiento de clientes con la finalidad de elaborar planes estratégicos. Nuestra investigación también está orientada a la clasificación de clientes, sin embargo, a diferencia de lo visto en la literatura nuestro trabajo está orientado a la elaboración de planeas tácticos por tratarse de actividades relacionadas con el transporte y distribución.

4.2 IMPLEMENTACIÓN DE LA FASE 1: ANÁLISIS DE DEMANDA

La complejidad de aplicar algún método de agrupamiento la encontramos en la disponibilidad, tipo, dimensión, legibilidad y entendimiento de la información a procesar.

Para desarrollar esta fase, es necesario la colaboración con la empresa, porque se convertirá en nuestra fuente de información primaria. El análisis de la demanda consiste entonces en diversos aspectos: familiarización y obtención de la información,

limpieza de datos y determinación de las variables a tratar en el estudio basado en el comportamiento histórico de las observaciones (clientes).

Familiarización y obtención de la información: Antes de proceder con el manejo de datos duros, se tuvieron un par de sesiones para conocer el significado de las bases de datos, su funcionalidad y procedencia, también se definió el formato en el que la empresa nos compartió su información.

Limpieza de datos: A fin de quedarnos con la información que para efectos de la investigación resulte de interés, la limpieza de datos consiste en: eliminar duplicados, rellenar espacios vacíos, eliminar registros, modificar filas o columnas, completar registros, agregar nueva información, filtrar y cruzar datos y extraer información adicional a partir de la actual. Este proceso requiere de tiempo, paciencia y las herramientas informáticas adecuadas. Tratando una base de datos cercana al millón de registros se utilizaron la herramienta *SQL* y *power BI*, que en combinación nos permitió llevar a cabo la limpieza de los datos y continuar con el análisis de la demanda.

En esta fase se manejó una base de datos de 42 columnas y 630,248 filas o registros, exportada de SAP a una hoja de cálculo, posteriormente trasladada a *SQLserver* (gestor de base de datos), para manipular eficientemente la información.

4.2.1 EXTRACCIÓN DE CARACTERÍSTICAS

Las variables a utilizar en el análisis clúster están basadas en la literatura y en la información disponible, de manera que para aplicar un algoritmo de agrupamiento se debe contar cuando menos con una variable, se decidió incluir aquellas variables que por coincidencia y recomendación estén presentes en nuestro análisis de la demanda. En la tabla 4.2, se muestran las variables utilizadas por distintos autores y las variables disponibles. Es importante resaltar que la distancia y el tiempo de viaje no son incluidos porque no buscamos diseñar clústeres balanceados en número

y forma geométrica. Por lo tanto, las variables que formarán parte del análisis de clúster quedan indicadas en la tabla siguiente.

Tabla 4.2: Variables de agrupamiento

Autor (es)	Vol. de compra	Demanda/ transporte	Costo	Tiempo de viaje	Distancia	Accesibilidad
Bosque (1997)	✓		✓	✓	✓	✓
Salado (2004)			✓	✓	✓	✓
Dondo y Cerdá (2007)		✓	✓		✓	✓
Alañon (2008)	✓		✓			✓
Liao y Guo (2008)	✓	✓	✓	✓		
Obregón et al (2014)	✓	✓	✓		✓	✓
Ríos (2017)			✓	✓	✓	

Fuente: Elaboración propia basada en la literatura

Hemos decidido tomar las variables sugeridas en la literatura y que se encuentran disponibles en nuestro caso práctico (Volumen de demanda, demanda por clase de vehículo, costo, accesibilidad, consolidación).

Basado en la tabla anterior, a continuación, se muestran tabuladas las características (variables) —ver tabla 4.3 y 4.4— que alimentaran nuestro algoritmo de agrupamiento y que son producto de las fases anteriores.

Tabla 4.3: Estructura tabular (a)

Observación	Demanda	Consolidado	Costo	Total	Accesibilidad
1	-	-	-	-	-
2	-	-	-	-	-
.	-	-	-	-	-
.	-	-	-	-	-
n	-	-	-	-	-

Fuente: Elaboración propia

Tabla 4.4: Estructura tabular (b)

Vehículo clase 1	Vehículo clase 2	Vehículo clase 3	Vehículo clase 4	Vehículo clase 5
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

Fuente: Elaboración propia

Se cuenta con clientes que no permiten que sus órdenes sean entregadas en la misma unidad de carga, es decir el cliente A, no permite que la mercancía solicitada viaje con la mercancía solicitada del cliente B. Por tal motivo, en la tabla anterior observamos la inclusión de la característica consolidación por ser un aspecto de alta importancia para nuestro caso de estudio en la selección de la configuración del transporte y para actividades distribución.

Es común contar con variables de diferentes dimensiones y unidades, por esta razón para evitar la presencia de valores atípicos en el análisis clúster, se recomienda normalizar las variables a fin ajustar los valores medidos en diferentes escalas respecto a una escala común (ver ecuación 4.1). Para estandarizar nuestras variables haremos uso de la siguiente formulación estadística:

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (4.1)$$

donde:

Z_i : valor del dato i normalizado

x_i : conjunto de datos, $i=1,2,3,\dots,n$

μ : valor promedio de los datos x_i para cada variable

σ : desviación estándar de la muestra de datos x_i para cada variable.

Realizada la normalización de los datos continuamos con la implementación de la siguiente fase de la metodología.

4.3 RESULTADOS DE LA FASE 2: SELECCIÓN DEL ALGORITMO DE AGRUPAMIENTO

Para la selección de la técnica de agrupamiento se tomaron en cuenta las siguientes características: tipo de información existente, número de observaciones a incluir, finalidad del análisis de agrupación y la escalabilidad. Además, se tomaron en cuenta los métodos más empleados en la literatura en cuanto a la efectividad de los resultados, en la figura 4.2, se muestran los métodos de mayor incidencia para resolver problemas de agrupamiento.

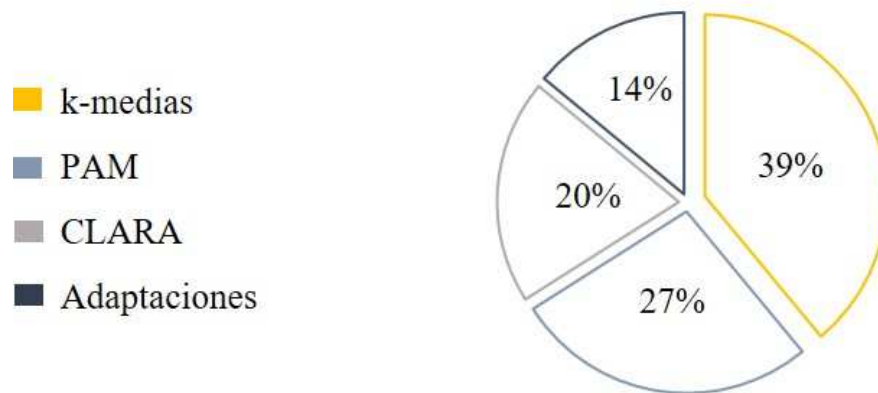


Figura 4.2: Métodos de agrupamiento

Fuente: Elaboración propia basada en la literatura.

En la revisión de literatura encontramos que el 39 % de los problemas de agrupamiento son abordados con la heurística *k-medias* por ser uno de los algoritmos de agrupación más simples y más empleados. De acuerdo a la consulta bibliográfica se decidió aplicar dicho método para abordar nuestro problema con la finalidad de encontrar un resultado factible que se adapte al objetivo de la investigación o bien nos permita validar la efectividad del método.

4.3.1 APLICACIÓN DEL ALGORITMO

La ejecución del algoritmo se codificó en el programa R studio por su disponibilidad como herramienta informática, visualización gráfica, recomendación en la literatura y por mantener una entendible interfaz.

Antes de ejecutar el algoritmo k -medias como se indicó en el capítulo 2, es indispensable determinar el número de agrupaciones (clases) apropiado para formar clústeres de clientes. Para llevar a cabo esta actividad es necesario realizar pruebas con 2, 3, 4 hasta k clústeres a fin de encontrar el número de grupos cuya varianza dentro de grupos sea la menor y la varianza entre grupos sea la mayor posible. En la tabla 4.5, podemos observar la cuantificación de la varianza intra-clúster e inter-clúster para 10 escenarios de prueba que van de 2 a 10 grupos, en la repetición realizada con 7 grupos se aprecia analíticamente que la varianza dentro de grupos alcanza el valor mínimo y la varianza entre grupos el máximo, después de la prueba con 7 grupos las repeticiones siguientes se comienzan a estabilizar, indicando que utilizar 7 grupos para el algoritmo k -medias es la mejor opción.

Tabla 4.5: Evolución de varianzas

Varianza/clase	2	3	4	5	6	7	8	9	10
Intraclase	29.041	28.634	27.886	27.110	26.794	25.201	26.229	25.790	25.377
Interclase	2.016	2.424	3.172	3.948	4.263	4.857	4.828	5.268	5.680

Fuente: Elaboración propia utilizando resultados de RStudio

A través del diagrama de codo, donde visualmente se pronuncie un quiebre, también es posible determinar el número de clústeres o clases adecuado para clasificar un conjunto de clientes. En la figura 4.3, podemos visualizar el comportamiento de las varianzas intraclase e interclase que el diagrama nos ofrece. Tanto en el resultado analítico como en el gráfico, a partir de 7 grupos las varianzas se estabilizan, de tal forma que al agregar un grupo extra no se muestra diferencia significativa al trabajar con una cantidad de agrupaciones superior a la indicada y apreciada por el codo.

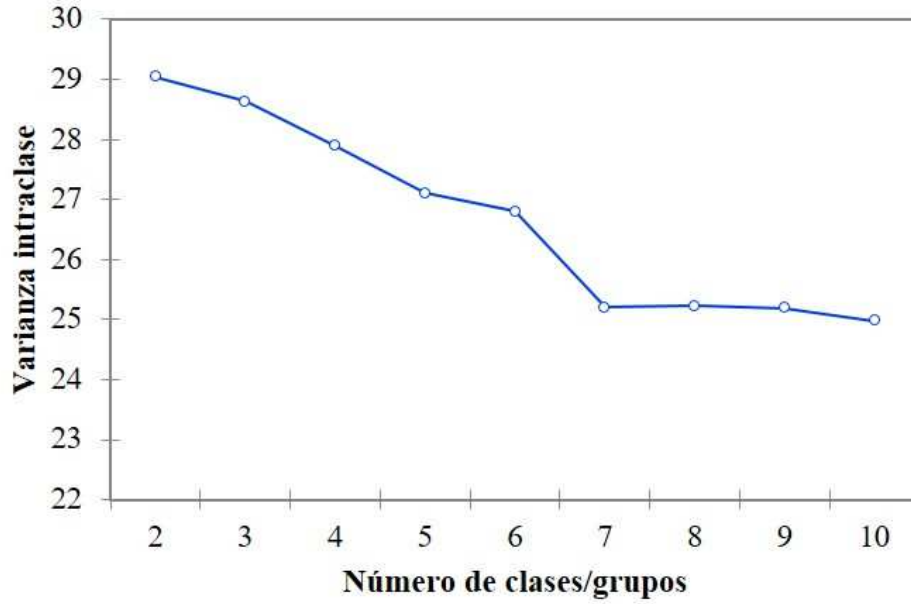


Figura 4.3: Evolución de varianzas

Fuente: Elaboración propia basada utilizando RStudio.

Una vez determinadas las variables y el número de grupos procedemos a la aplicación del método k -medias el cual es descrito mediante las siguientes expresiones.

Este algoritmo minimiza la suma de los cuadrados de los errores entre los puntos x_j y el correspondiente centroide c_i .

$$\min \sum_{i=1}^k \sum_{x_j \in c_i} \|x_j - \mu_i\|^2 \quad (4.2)$$

donde μ es la media de los puntos en c_i .

La asociación del elemento $x_j \in X$ con el centroide más cercano c_i es dada si:

$$S_i = \|x_j - c_i\| \leq \|x_j - c_k\| \quad (4.3)$$

donde cada x_j va exactamente dentro de un c_i

Los k centros iniciales se seleccionan aleatoriamente del conjunto de elementos X y son recalculados en cada iteración a partir del valor medio de todos los elementos asignados a ese grupo:

$$c_i = \frac{1}{c_i} \sum_{x_j \in c_i} x_j \quad (4.4)$$

Los resultados del algoritmo de agrupamiento son graficados en la figura 4.4, dando lugar a la formación de clústeres, que representan aquellos clientes que por la medida de similaridad y por los criterios de agrupamiento previamente identificados son similares entre si y disjuntos del resto de las agrupaciones.

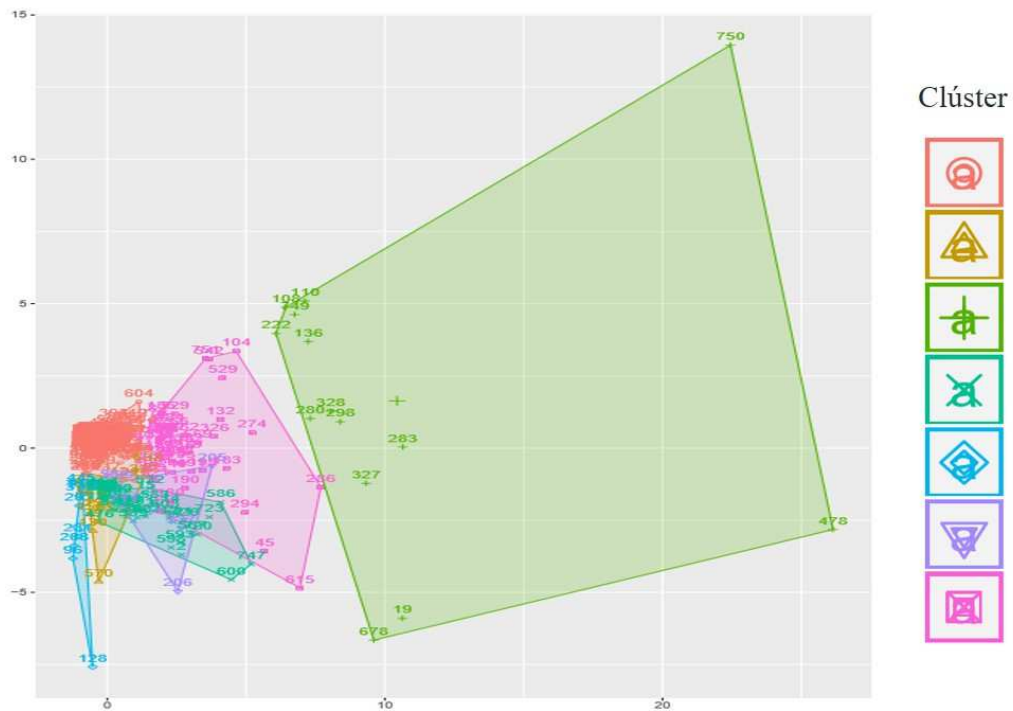


Figura 4.4: Clases (clústeres) resultantes

Fuente: Elaboración propia basada utilizando RStudio.

Analíticamente, obtenemos las medidas para cada clúster que indican el nivel de cohesión entre grupos (ver tabla 4.6). Cada agrupación obtenida tiene una varianza distinta, por lo que, a mayor sea la dispersión significa que los elementos (clientes)

comparten menos características, mientras que una agrupación con menor varianza representa a clientes afines en cuanto a volumen de compra, demanda de transporte, costo y accesibilidad. Se detectó por medio de cruzamiento y análisis de base de datos, que la alta variabilidad presente en el clúster número tres se debe a la naturaleza de los clientes agrupados, esto significa que la entrega de mercancía es realizada a centros de distribución, almacenes y puntos de *cross-dock* de la misma empresa, donde la proporción entre el costo del flete es menor que la demanda únicamente para este tipo de clientes, y que visto desde el nivel de accesibilidad, lo encontramos por encima de 1.

Por el contrario, si dirigimos la atención al resultado gráfico o al analítico, podemos observar que el clúster número uno, es la agrupación con mayor afinidad entre clientes. Esto se debe la fuerte relación que existe entre el costo, volumen de demanda y demanda de transporte y a la presencia de clientes que permiten viajes consolidados.

Tabla 4.6: Agrupamiento de los clientes según k -medias

Clase	1	2	3	4	5	6	7
Varianza dentro del grupo k	0.32	0.30	59.36	18.90	7.70	3.47	23.42
Distancia mínima al centroide	0.06	0.16	2.82	0.82	0.47	1.20	1.34
Distancia media al centroide	0.29	0.48	5.89	3.59	2.02	1.73	3.72
Distancia máxima al centroide	4.30	0.99	25.21	11.25	8.20	2.85	24.48

Fuente: Elaboración propia utilizando resultados de RStudio.

Para los clústeres 2,4,5,6,7 se sigue la misma dinámica de interpretación., varianzas elevadas indican dispersión entre las observaciones (clientes) de un clúster y varianzas cercanas a 0, señalan alto nivel de afinidad.

4.4 IMPLEMENTACIÓN DE LA FASE 3: VALIDACIÓN DE RESULTADOS

Como se mencionó con anterioridad, el algoritmo de agrupamiento k -medias es una técnica heurística que no garantiza el óptimo global para el problema de agrupamiento, sin embargo, puede ofrecer soluciones factibles que expliquen el comportamiento de los grupos mediante las varianzas. Para nuestro caso, el modelo empleado a través de las varianzas explica el 83 % de nuestros clústeres resultantes, dato que se obtiene automáticamente al codificar k -medias.

Cabe resaltar que, para validar nuestro modelo de agrupamiento utilizamos una función alternativa a los índices de validación vistos en la sección, se utilizó el valor R^2 arrojado por el modelo k -medias en el programa RStudio, el cual nos permitió evaluar la calidad general de las agrupaciones finales. Por lo tanto, al obtenerse el 83 % como medida de validación, podemos determinar que nuestros clústeres son aceptables.

En la tabla 4.5 de la sección anterior, debemos recordar que para determinar el número adecuado de clases fue necesario analizar la evolución de las varianzas de 2 a 10 clústeres, por lo que, la varianza dentro y fuera de los grupos finales resultó ser aquella donde la diferencia de cuadrados medio del error se estabiliza, concluyendo que los valores de las varianzas con ese número de clases validan nuestro modelo, ya que se minimiza la distancia intra-clúster e inter-clúster.

4.5 IMPLEMENTACIÓN DE LA FASE 4: DISEÑO DE CLÚSTERES

Recordando la implementación de la fase 2, gráficamente se lograba apreciar empalmes entre grupos, con la finalidad de no mostrar los empalmes se procedió a

descartarlos de la caracterización de los resultados a fin de obtener la figura siguiente (ver figura 4.5).

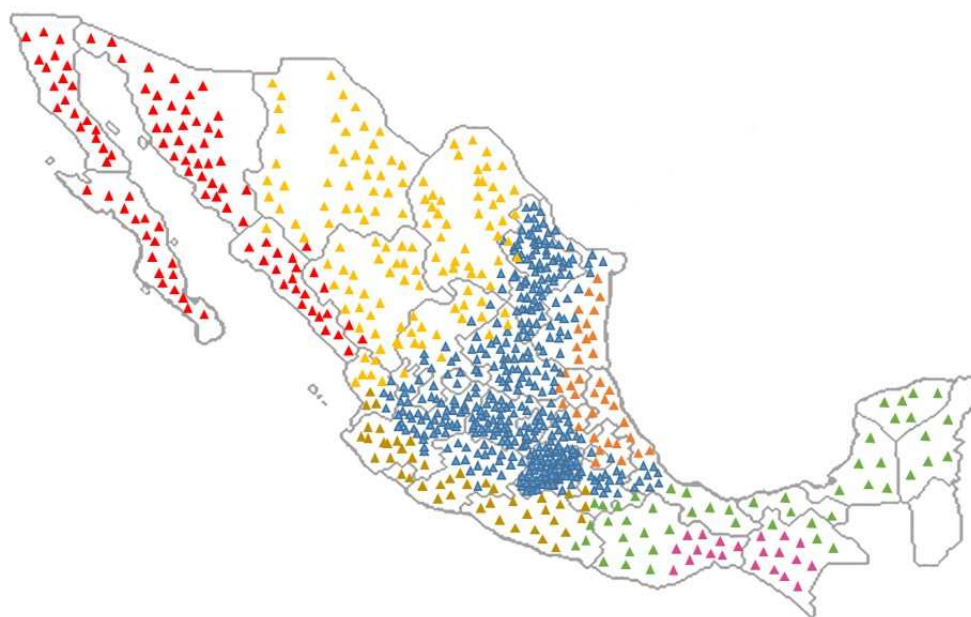


Figura 4.5: Formación de regiones

Fuente: Elaboración propia.

Es importante mencionar que la finalidad de esta fase es llevar la interpretación gráfica y analítica hacia un entorno geográficamente establecido, a fin de identificar las posibles regiones basadas en criterios técnicos.

4.6 IMPLEMENTACIÓN DE LA FASE 5: PREDICCIÓN

Como última fase de la metodología propuesta tenemos la predicción, una actividad de alta importancia para alcanzar los objetivos del presente trabajo de investigación, la cual consiste en encontrar el modelo de predicción apropiado para cada grupo obtenido de la fase anterior con la finalidad de cuantificar la demanda de transporte de cada conjunto de clientes, así como determinar la clase de vehículo a subcontratar.

La implementación comenzó con la manipulación de la información, esta vez para analizar los datos históricos que compondrán las series de tiempo, donde la única variable de interés será la demanda de las cinco clases de vehículos utilizados por la empresa para la distribución de productos finales. Para el manejo de la información se diseñó un el panel de control general elaborado para facilitar y ahorrar tiempo computacional (ver figura 4.6), el cual nos permitió trabajar con la información existente en la base de datos original por medio de herramientas informáticas *power BI*.



Figura 4.6: Panel de análisis para base de datos

Fuente: Elaboración propia utilizando *Power BI*.

Con la herramienta diseñada fue posible conocer el comportamiento de las series de tiempo en cada grupo constituido para determinado periodo, vehículo y origen-destino.

Tras haber analizado el comportamiento general de la variable dependiente o predictora (demanda del transporte), se encontró la presencia de características de nivel, tendencia y estacionalidad (ver figura 4.7). Por lo que, los modelos susceptibles a evaluar corresponden a metodos por descomposición, Holt Winters y modelos *ARIMA*.

Con el objetivo de seleccionar el mejor pronóstico, mediremos el error hacien-

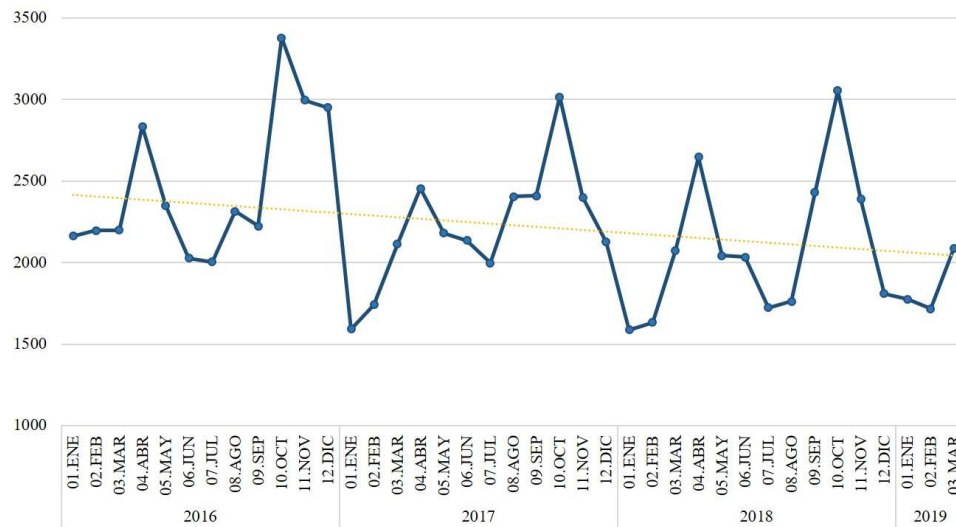


Figura 4.7: Demanda histórica acumulada

Fuente: Elaboración propia.

do uso de distintas medidas de precisión (MSE , $MAPE$, MAD y $RMSE$) para los metodos bajo experimentación.

4.6.1 DETERMINACIÓN DEL ESQUEMA DE LA SERIE TEMPORAL

Antes de ejecutar el modelo predictivo por descomposición y Holt-Winters, es necesario conocer el tipo de esquema que guardan nuestra serie de tiempo, puede tratarse de un esquema aditivo o multiplicativo. No olvidemos que para determinar el esquema de los datos es necesario calcular la media y la desviación estándar del mismo periodo a lo largo de la historia. En la tabla 4.7, se demuestra por medio de estadística descriptiva que tipo de esquema obedece. Este tipo de análisis es utilizado para modelos por descomposición y modelos de alisamiento exponencial de dos y de tres parámetros. Por lo que, al trabajar con una serie de tiempo con tendencia y estacionalidad se requiere conocer de qué tipo de esquema se trata (ver figura4.8).

Tabla 4.7: Análisis descriptivo del esquema temporal

Mes	promedio	Desv. estándar
Enero	1670.667	221.404
Febrero	1753.667	199.650
Marzo	2012.000	35.506
Abril	2534.000	129.066
Mayo	2078.667	32.356
Junio	1968.000	89.878
Julio	1822.667	89.078
Agosto	2062.000	238.512
Septiembre	2258.667	159.992
Octubre	2943.667	105.648
Noviembre	2489.667	227.311
Diciembre	2206.667	417.498

Fuente: Elaboración propia basada en el caso de estudio.

Nota: Cuando las medidas típicas de cada periodo crece a medida que crece el valor medio, es indicativo de que el patrón de agregación de las componentes es multiplicativo, en caso contrario se utiliza el esquema aditivo.

Se puede apreciar una nube de puntos dispersa que no permite distinguir la tendencia y, como se demostró en la tabla anterior pudimos observar que no hay crecimiento proporcional entre la demanda de transporte promedio y su desviación típica, por lo tanto estamos frente a un esquema aditivo.

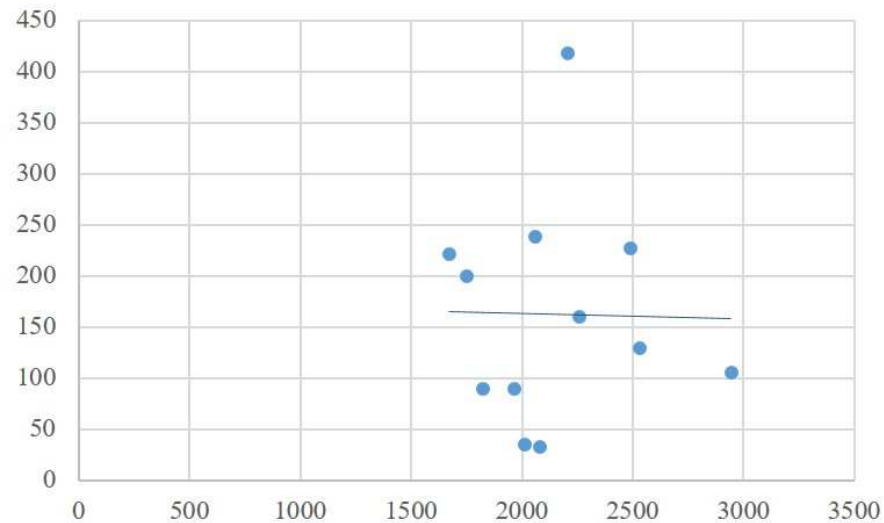


Figura 4.8: Tipo esquema para la serie temporal

Fuente: Elaboración propia basada en el caso de estudio

4.7 EXPERIMENTACIÓN CON MODELOS DE PREDICCIÓN

A continuación mostramos la implementación de los modelos predictivos que se adaptaron a nuestro caso de estudio y a las características de los datos.

4.7.1 MODELO POR DESCOMPOSICIÓN

En primer lugar, aplicamos el método por descomposición bajo esquema aditivo haciendo uso de la herramienta informática para uso estadístico Minitab 18. Determinamos las medidas de precisión empleando las formulaciones matemáticas (sección 2.4.3) utilizando hojas de cálculo para con la finalidad de verificar los resultados entregados por Minitab 18. En la figura 4.9, observamos el comportamiento presentado por este modelo respecto de la información real. Los valores ajustados y las estimaciones se muestran en la lista de apéndices.

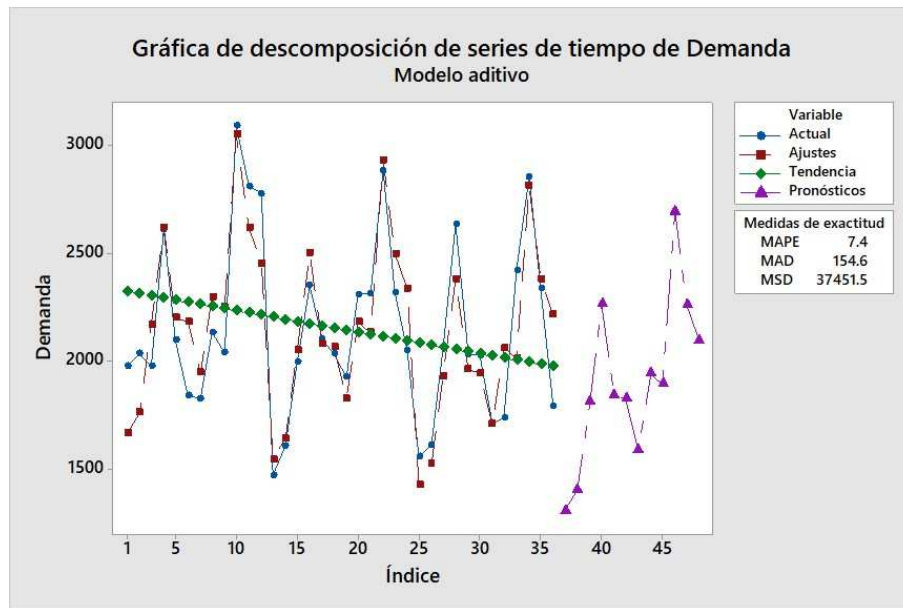


Figura 4.9: Modelo por descomposición

Fuente: Elaboración propia utilizando Minitab 18

4.7.2 EXPERIMENTACIÓN CON EL MODELO ARIMA

Esta clase modelos se conocen con el nombre genérico de ARIMA (Autoregressive Integrated Moving Average) que deriva de sus tres componentes AR (autorregresivo), I (integrado) y MA (medias móviles). Estos modelos son ampliamente recomendados cuando se tienen datos históricos como mínimo de 50 observaciones en la serie temporal, sin embargo, en nuestro caso práctico contamos con 44 observaciones, por lo que, hemos decidido aplicar ARIMA con este número de observaciones a fin de medir el desempeño que ofrecen estos metodos de carácter estocástico respecto al comportamiento real de la demanda de transporte.

Mediante programas de cómputo especializados (Eviews 10), medimos la precisión de los modelos ARIMA sugeridos por la herramienta informática. Se evaluaron 395 modelos de orden (p, d, q) (P, D, Q) que más se adaptaron a nuestra serie de tiempo (ver figura 4.10).

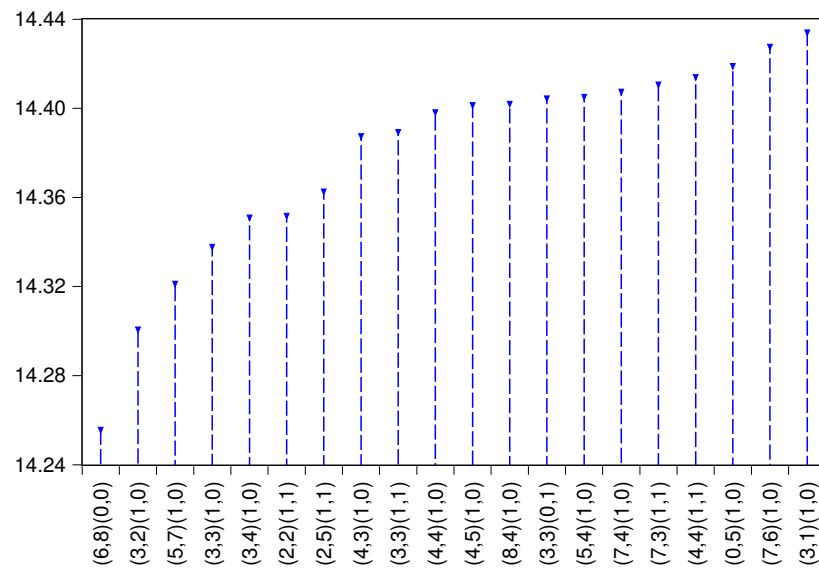


Figura 4.10: Modelos ARIMA evaluados

Fuente: Elaboración propia utilizando Eviews 10

Para determinar el modelo autorregresivos que representara a este método de pronóstico, evaluamos los modelos sugeridos utilizando tres medidas de precisión (ver tabla 4.8).

Tabla 4.8: Modelos ARIMA (p,d,q)(P,D,Q)

Modelo	RMSE	MAPE	MAD
(3,0,2) (1,0,0)	292.461	9.720	207.886
(3,0,2) (1,0,1)	260.703	8.522	167.460
(3,0,3) (0,1,1)	296.552	9.707	206.866
(2,0,2) (0,1,1)	248.181	8.692	170.053
(4,0,3) (0,1,1)	246.645	7.752	146.266
(3,0,4) (0,1,1)	246.883	7.776	147.788
(3,0,0) (1,0,0)	304.084	10.420	226.028
(5,0,3) (0,1,1)	253.690	7.696	146.681
(6,0,3)(0,1,1)	256.958	7.324	139.092
(6,0,3)(1,1,0)	261.782	7.398	140.902
(7,0,5)(0,1,1)	298.308	7.213	140.032
(8,0,6) (0,1,1)	322.382	6.966	133.529
(6,0,8) (0,1,1)	186.189	6.776	129.816
(7,0,5)(0,1,1)	290.773	6.921	135.396
(7,0,5)(1,1,0)	289.667	7.076	135.413

Fuente: Elaboración propia utilizando Eviews 10.

El modelo que redujo al máximo la variabilidad fue el de orden (6,0,8) (0,1,1), por consiguiente, será comparado contra el modelo por descomposición y Holt Winters.

4.7.3 EXPERIMENTACIÓN CON EL MODELO HOLT WINTERS

Por las características encontradas en nuestro análisis de la demanda de transporte correspondiente a esta fase (predicción), deducimos que el método para pronosticar de 3 parámetros (Holt Winters) es posiblemente el más apropiado de imple-

mentar. Este método de predicción lo ejecutamos bajo el esquema aditivo (recordar sección 4.6.1) y fue ejecutado en las herramientas (SPSS, Minitab y Eviews 10). Encontramos que este método es superior a los modelos por descomposición y modelos *ARIMA*. Es por tanto, que describimos matemáticamente el modelo a implementar. Se incorporan 3 constantes, que son α , β y γ , representan en una serie de tiempo el nivel, la tendencia y la estacionalidad respectivamente.

$$L_t = \alpha \frac{Y_t}{S_{T-P}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.5)$$

Por su parte, la ecuación para obtener la tendencia, T_t , es:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.6)$$

La estacionalidad S_t se obtiene con la ecuación siguiente:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-p} \quad (4.7)$$

Finalmente, se obtiene el pronóstico suavizado \hat{Y}_t :

$$\hat{Y}_t = (L_{t-1} + T_{t-1})S_{t-p} \quad (4.8)$$

Para estimar los factores estacionales de un periodo anterior, el procedimiento es el siguiente:

Primero debe estimarse T_0 :

$$T_0 = \frac{\overline{D_2} - \overline{D_1}}{n} \quad (4.9)$$

Luego se obtiene el valor L_0 :

$$L_0 = \overline{D_1} - \frac{n+1}{2}T_0 \quad (4.10)$$

Cada L siguiente se incrementa en el valor inicial de la tendencia T_0 :

$$L_t = L_0 + nT_0 \quad (4.11)$$

Se obtienen los factores estacionales para periodos posteriores:

$$S_t = \frac{D_t}{L_t} \quad (4.12)$$

donde:

L_t : es el nivel en el periodo t , α es el parámetro de nivel

T_t : es la tendencia en el periodo t , γ es el parámetro de tendencia

S_t : es el componente estacional para el periodo t , β es el parámetro estacional

p : indica el periodo estacional

Y_t : es el dato real del periodo t

\hat{Y}_t : valor estimado un periodo adelante en el tiempo t

L_{t-1} : es el factor de nivel del periodo anterior

T_{t-1} : es el factor de tendencia del periodo anterior

S_{t-p} : es el factor estacional del periodo anterior

T_0 : representa la tendencia para el periodo inicial

L_0 : representa el parámetro de nivel en el periodo inicial

n : número de estaciones anuales.

El modelo, además de probarlo con las herramientas informáticas mencionadas, también se desarrolló en el entorno de macros de Microsoft Excel empleando las formulaciones anteriores. Los valores ajustados obtenidos tras implementar dicho modelo los podemos visualizar en figura 4.11. Los valores óptimos de los parámetros

fueron: $\alpha=0.07$, $\beta=0.01$ y $\gamma=0.1$, los cuales obtuvimos en las herramientas informáticas y en el elaborado en MS Excel.

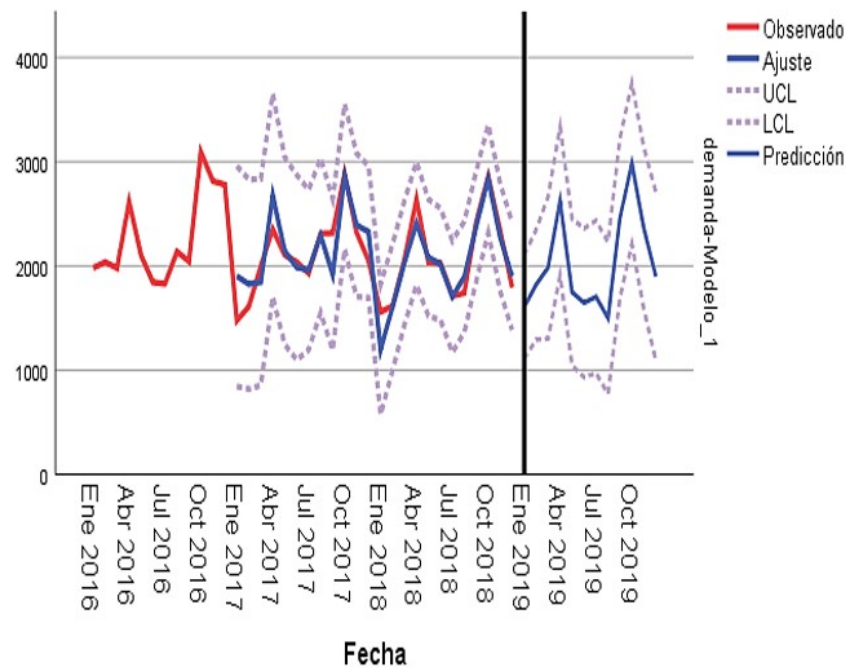


Figura 4.11: Modelo Holt Winters ajustado

Fuente: Elaboración propia utilizando SPSS

Cabe resaltar que, el modelo de predicción de alisamiento triple es recomendado utilizarse para realizar proyecciones al corto y mediano plazo, es decir, para estimar periodos mínimos de un mes y no mayores a dos años, de lo contrario la herramienta predictiva puede generar inconsistencias en los resultados (alta variabilidad).

4.7.4 ANÁLISIS COMPARATIVO

Para seleccionar el modelo predictivo que mejor explique la serie de tiempo bajo análisis, recurrimos al cálculo de las medidas de precisión descritas en la sección 2.4.3. Con la finalidad de recomendar la herramienta informática especializada cuya interfaz más comprensible, cuyo tiempo computacional sea inmediato, cuya disponi-

bilidad del programa esté al alcance, y cuyos resultados sean confiables, empleamos SPSS, E-Eviews y Minitab, como paquetes informáticos especializados. En la tabla 4.9, podemos apreciar el resumen en términos porcentuales tras la implementación de los métodos predictivos, las medidas de precisión y de las herramientas informáticas utilizadas.

Tabla 4.9: Análisis comparativo de los modelos predictivos evaluados

Método	Herramienta informática	<i>MAPE</i>	<i>MAD</i>	<i>MSE</i>	<i>RMSE</i>
Descomposición	Minitab SPSS Excel	7.4314	154.6057	37451.5395	193.5240
Holt Winters	Eviews SPSS Excel	6.7174	139.7569	30330.1935	174.1556
<i>ARIMA</i>	Eviews SPSS	6.7727	136.7510	34666.5123	186.1894

Fuente: Elaboración propia basada en la experimentación de la fase 5

De acuerdo a la literatura especializada, el principal indicador de exactitud de un estimador, sugerido y utilizado para la elección del mejor modelo de predicción, es el error cuadrático medio (*MSE*). Este mide la diferencia promedio entre el valor real y el estimado, ya que castiga aquellos periodos donde las diferencias (errores) son más grandes a comparación de otros, elevando los errores al cuadrado, tratándose de estimaciones por debajo o por encima del valor real y, que, al promediar los periodos comprendidos en una serie temporal, esta medida de precisión pondrá en tela de juicio la calidad de cualquier método de pronóstico para realizar proyecciones. Por otra parte, debemos considerar que esta medida de dispersión también representa una función de riesgo, es decir, existe la posibilidad de sobrestimar o desestimar una variable que esté altamente relacionada con los costos, recursos y nivel de servicio de cualquier empresa en caso de seleccionar el modelo predictivo equivocado.

El RMSE, indica las variaciones en unidades, el *MAD* nos indica las variaciones en términos reales, pero no penalizadas, tal como el *MSE*, y por último el *MAPE*, el cual expresa en términos porcentuales el nivel de precisión que entrega un modelo de pronóstico (recordar sección 2.4.3).

Por lo anterior, determinamos que, el modelo de predicción Holt Winters es

el apropiado para trabajar en la fase final de nuestra metodología concerniente a la predicción de la demanda de transporte, por número y por configuración vehicular.

4.7.5 EJECUCIÓN DEL MODELO HOLT WINTERS A LOS CLÚSTERES RESULTANTES

Debido a que la variable a pronosticar (demanda de transporte por clase de vehículo) presenta las mismas características de una serie de tiempo en todos los clústeres (estacionalidad, tendencia, y nivel), se aplicó el modelo de suavizado exponencial triple tras previo análisis de la demanda a nivel clúster en la herramienta mostrada anteriormente (volver a figura 4.6). De acuerdo a la experiencia con las herramientas informáticas, determinamos que el programa SPSS posee una interfaz más amigable, el tiempo computacional es inmediato, el programa se encuentra al alcance y los resultados arrojados son fiables y comprensibles por el usuario, por lo que, la implementación del modelo Holt Winters se llevó a cabo en dicho programa. Con la finalidad de evaluar la calidad de las estimaciones respecto de los valores reales, se decidió emplear el *MAPE* (Según la literatura consultada) por ser el métrico que proporciona una mejor interpretación y sencilla comprensión para el sector industrial (ver tabla 4.10). La tabla siguiente incorpora información por clúster, clase de vehículo y porcentajes de error.

Tabla 4.10: Análisis de error en el pronóstico a nivel clúster

Clúster	Modelo de predicción	Vehículo 1	Vehículo 2	Vehículo 3	Vehículo 4	Vehículo 5	\hat{x}
		<i>MAPE</i>	<i>MAPE</i>	<i>MAPE</i>	<i>MAPE</i>	<i>MAPE</i>	
1	Holt Winters	10.30 %	5.80 %	4.40 %	5.03 %	3.66 %	5.84 %
2	Holt Winters	5.17 %	3.71 %	5.33 %	3.33 %	4.72 %	4.45 %
3	Holt Winters	6.49 %	8.60 %	5.21 %	7.10 %	6.55 %	6.79 %
4	Holt Winters	4.40 %	4.10 %	5.69 %	5.52 %	5.27 %	5.00 %
5	Holt Winters	6.92 %	3.95 %	5.60 %	4.88 %	3.03 %	4.88 %
6	Holt Winters	9.56 %	3.24 %	3.77 %	6.00 %	4.46 %	5.41 %
7	Holt Winters	5.79 %	6.38 %	4.80 %	3.70 %	7.52 %	5.64 %
Promedio por vehículo		6.95 %	5.11 %	4.97 %	5.08 %	5.03 %	5.43 %

Fuente: Elaboración propia basada en la implementación de la fase 5

Es importante que notemos que el nivel de precisión es de gran calidad para cada clúster, así mismo el error promedio con el que proyectamos hacia un año es del 5.43 %, lo que significa que la combinación de técnicas de agrupamiento y pronósticos proporcionan una confiabilidad en las estimaciones del 94.57 %.

4.7.6 ANÁLISIS DE SENSIBILIDAD

Ha llegado el momento de la comprobar la efectividad de nuestro proyecto aplicado, para ello formularemos tres escenarios de decisión basados en el *MAPE* (ver tabla 4.11), el primero consiste en determinar el error porcentual promedio de cada clase de vehículo encontrado en cada región propuesta (clúster de clientes) empleando tecnicas de agrupamiento y modelos predictivos, el segundo únicamente se basa en el análisis y ejecución de pronósticos, por último, el tercer escenario, el cual muestra la situación actual de nuestro caso de estudio y que establece el punto de referencia de los escenarios anteriores.

Tabla 4.11: Análisis de sensibilidad basado en los tres escenarios

Método	Vehículo clase 1	Vehículo clase 2	Vehículo clase 3	Vehículo clase 4	Vehículo clase 5	\hat{x}
Análisis de clúster	6.95 %	5.11 %	4.97 %	5.08 %	5.03 %	5.43 %
Sin agrupamiento	16.12 %	10.45 %	12.62 %	11.78 %	14.34 %	13.06 %
Actual						18 %

Fuente: Elaboración propia basada las técnicas aplicadas

Por motivos de disponibilidad de información, en el escenario actual establecido por la empresa, solamente contamos con el error general, debido a la ausencia de indicadores que midan el error o nivel de precisión que sus habituales proyecciones a nivel vehículo (ver tabla 4.12). Es por tanto, que el 18 % representa el sesgo que puede presentar una estimación realizada por la empresa caso de estudio.

Tabla 4.12: Análisis de sensibilidad basado en los tres escenarios

Precisión en la planeación del transporte	
Técnicas de agrupamiento + Modelos Predictivos	94.57 %
Predicción con modelo apropiado	86.94 %
Cálculo habitual	82 %

Fuente: Elaboración propia basada en la confiabilidad de los métodos empleados

En la tabla anterior mostramos el nivel de precisión alcanzado bajo los tres escenarios, permitiéndonos comprobar nuestra hipótesis formulada en el capítulo 1, así como la funcionalidad brindada a la empresa caso de estudio para realizar el proceso de planeación enfocado al transporte.

4.8 CONCLUSIONES DEL CAPÍTULO

A partir del análisis de los resultados arrojados en las experimentaciones concluimos que, agrupar-primero, pronosticar-después, es una filosofía que arroja buenos resultados y es efectiva para los procesos de planeación presentes en logística y cadena de suministro.

Encontramos que las técnicas de agrupamiento nos permiten establecer control sobre el abanico de clientes, puesto que, los clasifica de acuerdo a criterios técnicos específicos con la finalidad de tomar de decisiones estratégicas, tácticas y operativas hacia grupos afines, permitiendo visualizar geográficamente las concentraciones de clientes mayoristas.

Tras analizar el comportamiento de la demanda para un periodo mayor a 5 años, pudimos encontrar que el modelo Holt-Winters posiblemente sea superado por modelos autorregresivos (ARIMA), esto se debe a que los modelos autorregresivos de media móvil son ampliamente recomendados por la literatura cuando se tiene series de tiempo mayores a 5 años, de tal forma que al haber trabajado con datos históricos a 3 años el modelo Holt-Winters entrego mejores resultados.

Es necesario que tengamos en cuenta que, esta combinación de herramientas trabaja de manera separadas y que al emplearlas con la secuencia propuesta generan distintos resultados. En el caso de añadir observaciones al abanico de clientes, el algoritmo d agrupamiento deberá ejecutarse nuevamente, incorporando las características consideradas. Por otra parte, si nuestro abanico de clientes permanece constante por determinado periodo, podremos generar proyecciones mínimo a un mes y máximo a dos años (utilizando Holt-Winters) para diferentes clases (clústeres) hasta que sean incorporados nuevos clientes.

Actualmente, nuestro caso de estudio mantiene un nivel de precisión en sus estimaciones del 82 %, por otra parte, implementando la metodología propuesta, la precisión aumenta al 94 %. Entonces, haciendo uso de ambas herramientas obtuvimos una mejora general en la estimación del número y tipo de caja del 12 %, y empleando únicamente modelos predictivos obtuvimos una mejora del 6 % en comparación de los métodos aplicados actualmente.

CAPÍTULO 5

CONCLUSIONES

Nuestra investigación propone un sistema para la gestión y planeación del transporte basado en herramientas de agrupamiento, minería de datos, extracción de características y diseño de clústeres, que en combinación buscan clasificar un conjunto de clientes con características y necesidades distintas en grupos de clientes por afinidad operativa, con la finalidad de brindar servicios de forma segmentada en tanto a esto refiera.

Posteriormente usamos un modelo matemático de predicción que, basado en los grupos previamente conformados y en la información histórica concerniente a las variables de interés (demanda de transporte y configuración del vehículo), genera proyecciones para cada clase de vehículo a nivel clúster, o bien, predice las variables de interés sin considerar agrupamientos, presentando un soporte para la toma de decisiones basado en técnicas cuantitativas.

5.1 CONCLUSIONES

Podemos concluir que la implementación de una herramienta que involucra el uso de técnicas de agrupamiento, complementada con la aplicación de un modelo matemático predictivo, entregan planeaciones efectivas que abordan directamente el

problema descrito en el capítulo 1, y aunado a la experiencia de los analistas en planeación del transporte y de los expertos en logística de nuestro caso de estudio, se fortalece tal actividad. Sin embargo, es una decisión estratégica que corresponde a nuestro caso de estudio, considerar segmentar de la forma propuesta todo el abanico de clientes o bien únicamente implementar la fase predictiva, cualquiera que sea su elección, quedó demostrado que en la implementación ya sea en conjunto o por separado de las herramientas abordadas, se obtiene en promedio una mejora del 10 %.

Emplear desde el inicio técnicas de clasificación, nos permitió identificar el comportamiento de los distintos clientes para determinar la importancia que cada uno representa y, por lo tanto, la atención que deberá ser brindada para cada grupo de clientes. Al conocer los hábitos de compra y sus condiciones de entrega, podemos llevar a cabo actividades de abastecimiento eficientes y enfocadas.

5.2 CONTRIBUCIONES

En este trabajo aplicamos un modelo de optimización para resolver el problema de agrupamiento motivado por un problema del mundo real, que a diferencia de los encontrados en la literatura, presenta un número de observaciones (clientes) alto (entre 1000 y 2000), cuenta con cientos de miles de registros para cada observación e incluye y hace uso de diversas variables de agrupamiento, orientadas a controlar la operación sobre el abanico de clientes y a la formulación de estrategias a nivel operativo, táctico o estratégico, de acuerdo a la misión y visión de la empresa. Con la finalidad de complementar el estudio y abordar directamente la necesidad de nuestro caso práctico, propusimos agregar e implementar la fase de predicción para dar uso específico a los clústeres de clientes conformados, la cual consistió en cuantificar la demanda de transporte esperada para cada clúster por medio de los métodos de pronóstico más indicados.

Nuestra contribución principal a la empresa caso de estudio fue entonces, crear una metodología basada en herramientas cuantitativas y una serie de recomendaciones técnicas para la elaboración de planes tácticos y toma de decisiones.

5.3 RECOMENDACIONES

Se sugiere adoptar programas informáticos especializados para la actualización constante del algoritmo de agrupamiento y para ejecución del modelo predictivo que mejor se ajuste a los datos actuales, contemplando gestores de base de datos capaces para sobrellevar la manipulación y análisis de grandes volúmenes de datos.

También se recomienda dividir la información en dos conjuntos. El primer conjunto le podemos denominar datos de control y utilizarlo para encontrar el mejor modelo de pronóstico. Y al segundo conjunto etiquetarlo como datos de predicción y no incluirlos en el análisis inicial.

5.4 POSIBLES LÍNEAS DE INVESTIGACIÓN A FUTURO

Concluida nuestra investigación, proponemos experimentar con otros algoritmos de agrupamiento empelando restricciones de tamaño y condicionando los criterios o variables.

A partir de la regionalización mostrada en la figura 4.5, proponemos reestructurar la red de distribución para cada clúster, de tal manera que la complejidad en el ruteo disminuya considerablemente al acortar el número de depósitos y nodos de visita, para ello será necesario realizar negociaciones con los proveedores del servicio de transporte para que se encarguen de cubrir las nuevas rutas de distribución.

También se recomienda hacer uso de las agrupaciones para fines mercadológicos, localizar instalaciones, establecer puntos de venta y porteo (*cross-dock*).

Por último, referente a las proyecciones y como trabajo a retomar, sugerimos externar la metodología propuesta hacia las diversas actividades logísticas que representan en promedio el 50 % del costo logístico total. Si está a consideración la incursión de otros modelos de predicción, no debemos dar por hecho que, a medida que aumente la complejidad del modelo para pronosticar encontraremos mejores resultados, más bien éste hecho dependerá de las fluctuaciones de la demanda a través del tiempo, donde existe la posibilidad de encontrar modelos tan sencillos de aplicar como promedio móvil o alisamiento simple, que puedan explicar efectivamente determinada serie de tiempo. Por lo anterior, invitamos a los interesados a retomar probables temas que pudiesen enriquecer la presente investigación.

BIBLIOGRAFÍA

- ADOLFSSON, A., M. ACKERMAN y N. C. BROWNSTEIN (2019), «To cluster, or not to cluster: An analysis of clusterability methods», *Pattern Recognition*, **88**, págs. 13–26.
- ALFARO, S. y C. AGUSTÍN (2017), «Descripción del procedimiento metodológico del análisis cluster no jerárquico con el algoritmo Clarans», .
- ANTÚN, J. P. (2005), *Logística de distribución física a minoristas*, UNAM.
- ARAUJO, R. (2015), «A semi-Supervised approach for kernel-based temporal clustering», .
- ARREOLA, R., L. MORENO y J. CARRILLO (2013), «Logística de transporte y su desarrollo», *revista Observatorio de la economía latinoamericana NISSN*, págs. 1696–8352.
- AZEVEDO, A. (2019), «Data mining and knowledge discovery in databases», en *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*, IGI Global, págs. 502–514.
- BALLOU, R. H. (2007), *Business logistics/supply chain management: planning, organizing, and controlling the supply chain*, Pearson Education India.
- BASSOLS-BATALLA, Á. (1993), «Formación de regiones económicas», *UNAM, Mexico*.

- BEN-DAVID, S. y M. ACKERMAN (2009), «Measures of clustering quality: A working set of axioms for clustering», en *Advances in neural information processing systems*, págs. 121–128.
- BID, P. T., BANCO INTERAMERICANO DE DESARROLLO (2013), «índices de gastos logísticos», <http://logisticsportal.iadb.org/node/4210>.
- BIOSCA, S., M. CHÁVEZ y E. B. QUEZADA (2014), «Road transport infrastructure and manufacturing location: an empirical evidence and comparative study between Tijuana and Nuevo Laredo, Mexico», *Frontera norte*, **26**(52), pág. 10.
- BOSQUE, J. S. (1994), «Sistema de información geográfica», *Estudios Geográficos*, **55**(214), pág. 201.
- CANACAR (2017), «Movimiento de carga por modo de transporte», <https://canacar.com.mx/stat/movimiento-carga-modo-transporte/>.
- CENDRERO, B. y S. TRUYOLS (2008), *El transporte: Aspectos y tipología*, Delta Publicaciones.
- CHEN, M.-S., J. HAN y P. S. YU (1996), «Data mining: an overview from a database perspective», *IEEE Transactions on Knowledge and data Engineering*, **8**(6), págs. 866–883.
- CHIAVENATO, I. y A. SAPIRO (2017), *Planeación estratégica*, McGraw-Hill Interamericana.
- CONABIO (2010), «Regiones económicas de México», <https://www.gob.mx/conabio>.
- CORDERO, H. *et al.* (1977), «Concentración industrial y poder económico en México», *Informe técnico*.
- CORREDERA, Y. D. (2012), «La logística empresarial y la administración de inventario», *Observatorio de la Economía Latinoamericana*, (168).

- DALTON, L., V. BALLARIN y M. BRUN (2009), «Clustering algorithms: on learning, validation, performance, and applications to genomics», *Current genomics*, **10**(6), págs. 430–445.
- DAVIS, H. *et al.* (2011), «Logistic Cost and Service annual global conference», *Council of Supply Chain Management Professionals*.
- DÍAZ, O., J. RUIZ y J. C. ZAVALA (2010), «Population pre-selection operators used for generating a non-random initial population to solve vehicle routing problem with time windows», *Scientific Research and Essays*, **5**(22), págs. 3529–3528.
- DONDO, R. y J. CERDÁ (2007), «A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows», *European Journal of Operational Research*, **176**(3), págs. 1478–1507.
- DRLIČIAK, M. y J. ČELKO (2016), «Implementation of transport data in to the transport forecasting in Slovakia», *Transportation Research Procedia*, **14**, págs. 1733–1742.
- ESCALERA, A. y M. G. BALDOQUÍN (2013), «Sistema soporte a la decisión para el agrupamiento de clientes de BRASCUBA SA», *Ingeniería Industrial*, **34**(2), págs. 143–154.
- FAYYAD, U. (1997), «Data mining and knowledge discovery in databases: implications for scientific databases», en *Proceedings. Ninth International Conference on Scientific and Statistical Database Management (Cat. No. 97TB100150)*, IEEE, págs. 2–11.
- FUNG, G. (2001), «A comprehensive overview of basic clustering algorithms», .
- GARCÍA, F., PEÑALVO (2017), «Revisión sistemática de literatura en los Trabajos de Final de Máster y en las Tesis Doctorales», .
- GEURS, K. T. y B. VAN WEE (2004), «Accessibility evaluation of land-use and transport strategies: review and research directions», *Journal of Transport geography*, **12**(2), págs. 127–140.

- GUERRERO, J. F., R. S. FERNÁNDEZ y J. C. ABAD (2006), «La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turístico», *Revista Europea de Dirección y Economía de la Empresa*, **15**(3), págs. 185–198.
- GUHA, S., R. RASTOGI y K. SHIM (1998), «CURE: an efficient clustering algorithm for large databases», en *ACM Sigmod Record*, tomo 27, ACM, págs. 73–84.
- GUTIÉRREZ, A. F. (2013), *Manual de pronósticos para la toma de decisiones*, Editorial Digital del Tecnológico de Monterrey.
- HALKIDI, M., Y. BATISTAKIS y M. VAZIRGIANNIS (2001), «On clustering validation techniques», *Journal of intelligent information systems*, **17**(2-3), págs. 107–145.
- HANKE, J. E. y D. W. WICHERN (2006), *Pronósticos en los negocios*, Pearson educación.
- HANSENS, D., L. PARSONS y R. SCHULTZ (2003), *Market response models: Econometric and time series analysis*, tomo 12, Springer Science & Business Media.
- HU, T.-L. y J.-B. SHEU (2003), «A fuzzy-based customer classification method for demand-responsive logistical distribution operations», *Fuzzy Sets and Systems*, **139**(2), págs. 431–450.
- HURTADO, B. A., J. M. ROBLES, J. M. PRECIADO y N. BAÑ (2018), «Logística de transporte y desarrollo local en organizaciones exportadoras de uva de mesa sonoreense», *Estudios sociales (Hermosillo, Son.)*, **28**(51), págs. 5–20.
- IMT, P. T., INSTITUTO MEXICANO DEL TRANSPORTE (2002), «Marco conceptual de la cadena de suministro: un nuevo enfoque logístico», .
- IZAR, J. (2007), *Modelos de pronóstico*, pág. 32.
- JAIN, A., N. MURTY y P. FLYNN (1999), «Data clustering: a review», *ACM computing surveys (CSUR)*, **31**(3), págs. 264–323.
- JINTANA, J. y T. MORI (2019), «Customer Clustering for A New Method of Marketing Strategy Support within the Courier Business», **31**(2), págs. 1–19.

- KANUNGO, T. *et al.* (2004), «A local search approximation algorithm for k-means clustering», *Computational Geometry*, **28**(2-3), págs. 89–112.
- LARRAÑAGA, P. y J. A. LOZANO (2001), *Estimation of distribution algorithms: A new tool for evolutionary computation*, tomo 2, Springer Science & Business Media.
- LIAO, K. y D. GUO (2008), «A Clustering-Based Approach to the Capacitated Facility Location Problem 1», *Transactions in GIS*, **12**(3), págs. 323–339.
- LIGHTNER, C., V. AGRAWAL, C. LIGHTNER y N. WAGNER (2016), «An evolutionary algorithm approach for the constrained multi-depot vehicle routing problem», *International Journal of Intelligent Computing and Cybernetics*, **9**(1), págs. 2–22.
- LIKAS, A., N. VLASSIS y J. J. VERBEEK (2003), «The global k-means clustering algorithm», *Pattern recognition*, **36**(2), págs. 451–461.
- LONG, D. (2006), *Logística internacional: administración de la cadena de abastecimiento global*, Editorial Limusa.
- LÓPEZ, R. (2018), «Regiones Socio-Económicas», <https://www.eweb.unex.es>, 23-08-19.
- LU, C. y L. KAO (2016), «A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server», *Engineering Applications of Artificial Intelligence*, **55**, págs. 231–238.
- MACQUEEN, J. *et al.* (1967), «Some methods for classification and analysis of multivariate observations», en *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, tomo 1, Oakland, CA, USA, págs. 281–297.
- MAKRI, M. y C. FOLKESSON (1999), «Accessibility measures for analyses of land use and travelling with geographical information systems», *Department of Technology and Society, Lund Institute of Technology, Sweden*, **1**.

- MARTÍNEZ, A. y M. A. CRUZ (2011), «Método de Agrupamiento no Supervisado para el Problema de Ruteo Vehicular con Restricciones de Capacidad en Vehículos», .
- MESA, R. y S. UKKUSURI (2015), «Demand clustering in freight logistics networks», *Transportation Research Part E: Logistics and Transportation Review*, **81**, págs. 36–51.
- MONTES, E., F. CALVETE y C. A. MANTILLA (2016), «Aplicación de series de tiempo en la realización de pronósticos de producción.», *Revista Fuentes, El Reventón Energético*, **14**(1).
- NALLUSAMY, R., K. DURAISWAMY, R. DHANALAKSMI y P. PARTHIBAN (2010), «Optimization of multiple vehicle routing problems using approximation algorithms», *arXiv preprint arXiv:1001.4197*.
- NGAI, E. W., L. XIU y D. C. CHAU (2009), «Application of data mining techniques in customer relationship management: A literature review and classification», *Expert systems with applications*, **36**(2), págs. 2592–2602.
- PACHECO, J. y O. VALENCIA (2005), «Análisis de nuevos métodos de clasificación. Un ejemplo ilustrativo de su uso en la agrupación de los municipios de Castilla y León», *Estudios de economía aplicada*, **23**(3), págs. 711–729.
- PAKYÜREK, M. *et al.* (2018), «Customer clustering using RFM analysis», en *2018 26th Signal Processing and Communications Applications Conference (SIU)*, IEEE, págs. 1–4.
- PAPA, E., C. SILVA, T. BRÖMMELSTROET y A. HULL (2015), «Accessibility instruments for planning practice: a review of European experiences», *Journal of Transport and Land Use*, **9**(3).
- PARDO, Á. y J. CAROD (2008), «Accesibilidad y localización industrial. Una aplicación a las regiones españolas fronterizas con Francia», *Revista de Estudios Regionales*, (82), págs. 71–103.

- PASCUAL, D., F. PLA y S. SÁNCHEZ (2007), «Algoritmos de agrupamiento», *Método Informáticos Avanzados*, págs. 164–174.
- PEREZ, R., S. SANCHEZ y J. BASTIDAS (2012), «Aplicación de modelos de pronósticos en productos de consumo masivo», *Biotecnología en el Sector Agropecuario y Agroindustrial: BSAA*, **10**(2), págs. 117–125.
- PINDYCK, R. S. y D. L. RUBINFELD (2001), «Econometría, modelos y pronósticos. Capítulo 8, Pronósticos con un modelo de regresión de una sola ecuación», *Editorial McGraw Hill. México, DF*, págs. 211–234.
- RAMÍREZ, A. C. (2009), *Manual de la gestión logística del transporte y distribución de mercancías*, Universidad del Norte.
- RENDÓN, T. A. y A. MORALES (2008), «Grupos económicos en la industria de alimentos: Las estrategias de Gruma», *Argumentos (México, DF)*, **21**(57), págs. 87–112.
- RÍOS, Q. G. y S. OBREGÓN (2017), «La accesibilidad de las autovías y la teoría de localización industrial», *Economía, sociedad y territorio*, **17**(55), págs. 581–617.
- RIVERA, R. A., L. M. DELGADO y J. DE JESÚS CARRILLO (2013), «Logística de transporte y su desarrollo», *Observatorio de la Economía Latinoamericana*, **185**.
- SALADO, G. M. J. (2004), «Localización de los equipamientos colectivos, accesibilidad y bienestar social», *Bosque Sendra, J. y Moreno Jiménez, A. Sistemas de información geográfica y localización de instalaciones y equipamientos. Madrid: Editorial RA-MA*, págs. 17–51.
- SALAZAR, M. A., R. RÍOS y J. GONZÁLEZ (2011), «A bi-objective programming model for designing compact and balanced territories in commercial districting», *Transportation Research Part C: Emerging Technologies*, **19**(5), págs. 885–895.
- SCT (2017), «NORMA OFICIAL MEXICANA NOM-012-SCT-2-2017», .

- SHAUB, D. (2019), «Fast and accurate yearly time series forecasting with forecast combinations», *International Journal of Forecasting*.
- SHEU, J.-B. (2007), «A hybrid fuzzy-optimization approach to customer grouping-based logistics distribution operations», *Applied Mathematical Modelling*, **31**(6), págs. 1048–1066.
- TAILLARD, É. D. (2003), «Heuristic Methods for Large Centroid Clustering Problems», *Journal of Heuristics*, **9**(1), págs. 51–73, URL <https://doi.org/10.1023/A:1021841728075>.
- TAO, C.-W. (2002), «Unsupervised fuzzy clustering with multi-center clusters», *Fuzzy Sets and Systems*, **128**(3), págs. 305–322.
- TIAN, K., J. LI, J. ZENG, A. EVANS y L. ZHANG (2019), «Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm», *Computers and Electronics in Agriculture*, **165**, pág. 104962.
- TORRES, M. (2011), «Pronósticos, una herramienta clave para la planeación de las empresas», *Instituto Tecnológico de Sonora [consultado 12 Sept 2019]. Disponible en: Disponible en: http://itson.mx/publicaciones/pacioli/Documents/no71/47a.pronosticos,_una_herramienta_clave_para_la_planeacion_de_las_empresas.pdf*.
- TORRES-RABELLO, R. (2012), «Precisión de Pronósticos», *Supply Chain Management*, págs. 1–2.
- VALLEJO HUANGA, D. (2016), «Clustering de Documentos con Restricciones de Tamaño», .
- VILLAGRA, A. y G. LEGUIZAMÓN (2007), «Metaheurísticas aplicadas a Clustering», .
- WILSON, J. H. y B. KEATING (1999), *Business Forecasting*, McGraw Hill, USA Transaction Volumes.

- YAN, M. (2005), *Methods of determining the number of clusters in a data set and a new clustering criterion*, Tesis Doctoral, Virginia Tech.
- YANG, M. y Y. NATALIANI (2017), «Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters», *Pattern Recognition*, **71**, págs. 45–59.
- YU, N., G. DE ROO, M. DE JONG y S. STORM (2016), «Does the expansion of a motorway network lead to economic agglomeration? Evidence from China», *Transport Policy*, **45**, págs. 218–227.
- ZHAO, C., M. JOHNSON y M. HE (2017), «Data mining with clustering algorithms to reduce packaging costs: A case study», *Packaging Technology and Science*, **30**(5), págs. 173–193.

RESUMEN AUTOBIOGRÁFICO

Alan Emilio Pahua Reyes

Candidato para obtener el grado de
Maestría en Logística y Cadena de Suministro

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:

AGRUPAMIENTO GEOGRÁFICO DE LA DEMANDA PARA LA
PLANEACIÓN DEL TRANSPORTE Y DISTRIBUCIÓN

Nacido en Páztcuaro Michoacán el 25 de agosto de 1992. Hijo de Juan Manuel Pahua Molina y de Maricela Reyes Ambris. Ingeniero Industrial egresado del Instituto Tecnológico de Morelia (ITM) en el año 2016.